Why Principals Often Give Overly High Ratings on Teacher Evaluations

Haim Shaked

Studies in Educational Evaluation (In press)

Abstract

Recent research found that principals who are required to evaluate their teachers often give higher ratings than what they think these teachers deserve**.** This study aimed to explore principals' considerations while evaluating teachers. Participants were 39 Israeli principals. Data were collected through semi-structured interviews and were analyzed in four stages – condensing, coding, categorizing, and theorizing. Four considerations emerged for principals' over-evaluations: (1) time constraints/prioritization (low perceived value for high time investment); (2) evaluation's ineffectiveness for improving teaching (via teacher development or dismissal); (3) the imprecision of teacher evaluation measurements; and (4) impingement on interpersonal relationships. This study demonstrated how principals serve as local mid-level policymakers by actively buffering, rather than bridging, the policies imposed on their schools from above.

Why Principals Often Give Overly High Ratings on Teacher Evaluations

Since the turn of the century, teacher evaluation has been introduced around the world with the goal of upgrading teacher functioning so as to raise the level of student learning (Tuytens & Devos, 2017). In fact, teacher evaluation serves as a key component in many countries' contemporary educational policies that aim to improve their school systems (Marzano & Toth, 2013; OECD, 2009). In particular, current accountability trends, which have made teachers individually accountable for student achievement to a greater extent than ever before, have increased the role of teacher evaluations in educational policies (Marchant, David, Rodgers, & German, 2015). Therefore, today's principals are required to evaluate their teachers regularly (Donaldson & Papay, 2015; Steinberg & Donaldson, 2016).

However, researchers found that principals very often give their teachers the highest possible ratings (Donaldson, 2009; Weisberg, Sexton, Mulhern, & Keeping, 2009), whereas teachers are rated as unsatisfactory only in rare cases (Kraft & Gilmour, 2017; Toch & Rothman, 2008). Such skewed ratings do not allow for discrimination between effective and ineffective teachers and do not provide high quality feedback to improve teachers' functioning (Marzano & Toth, 2013; Wechsler et al., 2007). Yet, the literature to date has barely investigated principals' core explanations for their over-evaluation of teachers (e.g., Kraft & Gilmour, 2017).

To further elucidate why principals frequently give higher ratings than they think teachers deserve, the current study aimed to explore Israeli principals' considerations during teacher evaluation. Thus, this study qualitatively examined a maximally differentiated sample of principals to elicit their perceptions and interpretations of the Israeli teacher evaluation policy, seeking to explain their reasons for inflating teachers' ratings.

Teacher Evaluation

Recently, teacher evaluation has become a preferred policy lever at the federal, state,

and local levels (Steinberg & Donaldson, 2016; Tuytens & Devos, 2017). In general, teacher

evaluation has two basic purposes: measuring teachers and developing teachers (Marzano,

2012). Teacher measurement discerns differences between various teachers' levels of

effectiveness, while teacher development provides teachers with meaningful feedback about

their practice in order to bring about improved instruction and achievements (Donaldson,

2009; Wechsler et al., 2007). To achieve these goals, principals from all over the world are

required to constantly evaluate their teachers (Marzano & Toth, 2013; OECD, 2009).

However, an increasing body of research has indicated that teacher evaluation by

principals actually fails to provide reliable information regarding teacher quality, because

teachers almost always receive high ratings from their principals. Toch and Rothman (2008)

discovered that 87% of the 600 schools in the Chicago school system did not rate even one

teacher as unsatisfactory even though 10% of those schools were classified as "failing

educationally" (Marzano & Toth, 2013, p. 2). The rating scale used in Chicago included four

grades: superior, excellent, satisfactory and unsatisfactory. Overall, only 0.3% of all

Chicago's 25,000 teachers were rated as unsatisfactory, while 93% of teachers in the system

were rated as "superior" or "excellent" (The New Teacher Project, 2007). Similarly, Weisberg

et al. (2009) found that in a district with almost 35,000 tenured teachers, only 0.4% were

given the lowest rating, while almost 70% received the highest. Weisberg et al. called this

phenomenon "the Widget Effect" (p.4):

> The Widget Effect describes the tendency of school districts to assume
>
> classroom effectiveness is the same from teacher to teacher. This decades-old
>
> fallacy fosters an environment in which teachers cease to be understood as
>
> individual professionals, but rather as interchangeable parts.

Weisberg and his colleagues (2009) further demonstrated that these inflated formal

teacher ratings did not reflect evaluators' actual ability to recognize differences in teachers'

effectiveness. They found that a high percentage of principals and teachers (81% and 57%, respectively) could identify a poor-performing teacher in their school, despite the fact that in most districts less than 1% of teachers were given an unsatisfactory rating. More recently, Kraft and Gilmour (2017) revisited these findings in 24 states that had adopted large-scale reforms in their teacher-evaluation methods. Although the full distribution of ratings was found to vary widely across states, with 0.7% to 28.7% of teachers rated as below proficient and 6% to 62% rated as above proficient, the percentage of teachers rated as unsatisfactory remained as before, at less than 1% in the vast majority of states.

In a preliminary attempt to qualitatively explore these quantitative findings, Kraft and Gilmour (2017) also interviewed 24 principals, who assigned their teachers an overall performance rating on a four-category rating scale based on their holistic assessment of evidence from various sources (performance measures based on standardized tests were not incorporated). Kraft and Gilmour's (2017) study yielded four possible reasons why so few teachers received below-proficient ratings, which often did not reflect principals' perceptions of teachers' actual performance. First, principals reported lacking the time needed to rate a teacher as unsatisfactory. Rating teachers as below proficient requires intensive amounts of time, which principals seldom have, to document their performance ("up to four unannounced formal observations," p. 241) and later to provide support by writing up and implementing improvement plans. Second, principals factored in teachers' potential and motivation when assigning an evaluation rating, especially when referring to teachers who were just beginning their careers. Third, principals wanted to avoid the personal discomfort involved in rating teachers as below proficient, particularly because such a rating might lead to these teachers' dismissal. Fourth, principals did not rate teachers as unsatisfactory due to their preference to avoid the long, laborious process of removing and replacing teachers. The current study sought to further elaborate on these initial interview findings, by conducting an in-depth

qualitative analysis of principals' considerations while evaluating their teachers. Moreover, replicating Kraft and Gilmour's (2017) study in various socio-cultural contexts may enable generalization of their findings to broader populations, possibly substantiating their international validity.

Teacher Evaluation vis-à-vis Instructional Leadership

Teacher evaluation has often been considered among the components of principals' instructional leadership (e.g., Hallinger & Wang, 2015; May & Supovitz, 2011; Stronge, Richard, & Catano, 2008; Supovitz, Sirinides, & May, 2010). In the instructional educational leadership approach, school principals engage in a wide range of activities aiming to promote their expected primary objective of explicitly improving the school's teaching and learning for all students (Goldring, Huff, May, & Camburn, 2008; Murphy, Neumerski, Goldring, Grissom, & Porter, 2016; Shaked, 2018; Walker & Slear, 2011). Despite considerable effort invested by researchers and policymakers in campaigns over the last 40 years aimed at framing instructional leadership as a key component of the principal's role (Prytula, Noonan, & Hellsten, 2013), recent studies have shown that the amount of time that most principals devote to actual activities aiming to improve their schools' teaching and learning has hardly changed (Goldring et al., 2015; Murphy et al., 2016). Although some principals do practice instructional leadership, many others do not (Camburn, Spillane, & Sebastian, 2010; Goldring et al., 2008; Grissom, Loeb, & Master, 2013; Horng, Klasik, & Loeb, 2010; May & Supovitz, 2011).

Several main barriers to progress have been mentioned in the literature in regard to principals' investment of time and effort toward instructional leadership activities, although these barriers did not specifically relate to the teacher evaluation component of instructional leadership. First, principals may lack sufficient time to engage in direct attempts to improve teaching and learning (Goldring et al., 2015; Murphy et al., 2016), largely because of ongoing

structural limits on their time, which pressure them to attend to other issues such as student affairs (Camburn et al., 2010). Moreover, while instructional leadership tasks require uninterrupted blocks of time for activities such as planning, writing, conferencing, observing, analyzing curriculum, and developing professional growth activities for staff, a principal's average workday is usually made up of a mosaic of activities, each of which is given brief attention only (Murphy et al., 2016; Prytula et al., 2013). Inasmuch as considerable time is spent on unplanned events and crisis solutions, principals' efforts to work on instructional matters seldom receive sufficient time resources during day-to-day school operations.

In addition to time constraints, many principals seem to lack the explicit knowledge-base and skill-set necessary to function as instructional leaders. Their "instructional leadership content knowledge" appears to be underdeveloped – referring to knowledge concerning how students learn specific subjects, which teaching methods are effective in which contexts, and the like (Goldring et al., 2015; Stein & Nelson, 2003). "Without an understanding of the knowledge necessary for teachers to teach well… school leaders will be unable to perform essential school improvement functions such as monitoring instruction and supporting teacher development" (Spillane & Louis, 2002, p. 97).

One of the capabilities needed to engage in effective instructional leadership is the capacity to build good relationships (Robinson, 2010). The influence of principals on students is mainly indirect (Murphy et al., 2016). Principals who enact instructional leadership do so by influencing teachers' teaching strategies and by increasing teachers' motivation, loyalty, satisfaction, and other factors that, in turn, influence student outcomes (Blase & Kirby, 2009; Louis, Dretzke, & Wahlstrom, 2010; Supovitz et al., 2010; Thoonen, Sleegers, Oort, & Peetsma, 2012). Through such positive relationships, instructional leaders can engage with teachers in productive and respectful conversations about the quality of teaching and learning (Le Fevre & Robinson, 2015). Indeed, positive principal-teacher relationships were shown to

help teachers adopt more effective teaching practices (Alsobaie, 2015), demonstrating a critical role in the improvement of student achievements (Edgerson, Kritsonis, & Herrington, 2006; Price, 2015; Price & Moolenaar, 2015; Tschannen-Moran & Gareis, 2015).

Uncovering School Principals' Policy Interpretation as Mid-Level Policymakers

Principals stand at the school "doorstep," as a mediating agent between the extra- and intra-school worlds (Kelchtermans, Piot, & Ballet, 2011; Maxcy, Sungtong, & Nguyen, 2010), negotiating between inside (within-school) desires and capacities and outside (national) demands and expectations (Louis & Robinson, 2012; Schechter & Shaked, 2017). Yet, rather than acting as mere gatekeepers (Salter, 2014), principals often act as unofficial mid-level policymakers who adjust and modify external policy to their particular school (Diamond, 2012; Louis & Robinson, 2012; Spillane & Kenney, 2012). By mediating between the school and external authorities, adapting and incorporating particular policy elements and practices to each school's specific characteristics, principals create new norms that change the original policy over time (Diamond, 2012; Louis & Robinson, 2012). In this unique position, principals determine the extent to which a school will adopt a "bridging" approach or will "buffer" against external influences (Kohansal, 2015; Paredes Scribner, 2013; Rutledge, Harris, & Ingle, 2010), depending on contextual and temporal issues related to policies' implementation (Hall & Hord, 2015; Honig & Hatch, 2004).

In the current case, where principals have been widely documented as deviating from the original intentions of national teacher evaluation planners by giving school teachers overly high ratings and avoiding low ratings (e.g., Toch & Rothman, 2008; Weisberg et al., 2009), qualitative inquiry into a heterogeneous sample of Israeli principals was undertaken to uncover possible patterns in principals' policy interpretation. Different principals may interpret the exact same policy idiosyncratically, varying in terms of content, focus, and intensity (Koyama, 2014; Levin & Datnow, 2012). They determine the extent to which their

schools will adopt a "bridging" vs. "buffering" approach towards the environment (Kohansal, 2015; Paredes Scribner, 2013; Rutledge, Harris, & Ingle, 2010). The current study aimed to elicit principals' unique voices that might reveal group trends explaining these mid-level policymakers' personal "mark" on policies received from above (Flessa, 2012; Kelchtermans et al., 2011).

Israeli Research Context

Israel's national school system serves about 1.6 million students (Israeli Central Bureau of Statistics, 2013). The Capstones Institute, which is responsible for preparing all school principals in Israel, has defined Israeli school principals' primary role as that of serving as instructional leaders in order to improve all students' education and learning (Capstones, 2008).

The New Horizon national reform was launched in the Israeli school system in 2009, introducing a new teacher evaluation system for elementary and middle schools. Under the new system, principals are to evaluate teachers who are candidates for tenure as well as teachers who are candidates for promotion to higher ranks. The new system calls for principals to formulate an evaluation following observation of three classes taught by the teacher, twice in a full classroom and once in a small group, which must include a conversation with the teacher before and after each observation. The principal is expected to use impressions from these classroom observations and general familiarity with the teacher's work as evidence to rate each of four domains of the teacher's performance on a 7-point scale from *Below basic level* to *Master*. The four rated domains are: identification with the role of teacher and commitment to the organization; proficiency in the subject matter and in methods and modes of teaching it; instruction quality, including components such as planning, teaching methods, and assessment; and participation in a professional community. Evaluated teachers who are candidates for tenure or promotion may appeal the principal's evaluation.

Performance measures based on student achievements are not incorporated into the evaluation. Improvement plans also are not part of the evaluation process (Israeli Ministry of Education, 2015).

Method

Qualitative interview methodology and content analysis were utilized in this study to provide rich textual descriptions of school principals' considerations in conducting teacher evaluations (Taylor, Bogdan, & DeVault, 2016).

Participants

Participants in this study were 39 elementary and middle school principals who had been asked to evaluate their teachers who are candidates for tenure or promotion to higher ranks using the new teacher evaluation system under the New Horizon national reform. Based on prior research findings regarding the high prevalence of over-evaluation of teachers (Donaldson, 2009; Kraft & Gilmour, 2017; Toch & Rothman, 2008), reviewed above, I did not look for principals whose schools showed signs of over-evaluation, but assumed that over-evaluation of teachers exists in many schools. The study was driven by the working assumption that quite a few principals would discuss their over-evaluation of teachers.

 Seeking to maximize the depth and richness of data, maximal differentiation sampling (Creswell, 2014), also known as heterogeneous sampling, was employed. This purposive sampling technique captures a wide range of perspectives, gaining significant insight into the phenomenon at hand by contemplating it from a variety of angles (Merriam, 2009). Maximal differentiation sampling was implemented in this study regarding principals' sex, age, years of experience, education, school level (elementary/middle), and school community's socioeconomic status. Participants were recruited through the assistance of regional superintendents. The study sample did not begin with a set number of participants, but rather developed on an ongoing basis as the study progressed (Taylor et al., 2016). Altogether, 72

school principals were approached, until 39 principals who could represent diverse sampling were obtained. Principals who did not want to participate explained their refusal mainly in lack of time.

Similar to national statistics (Israeli Central Bureau of Statistics, 2013), participating principals (12 males and 27 females) had 11 to 32 years of educational experience ($M =$ 21.17, $SD = 3.94$), which included one to 19 years of experience as principals ($M = 6.81$, $SD = 5.32$). The majority of the 39 interviewees ($n = 27$) held a master's degree, with 11 principals holding only a bachelor's degree and one principal with a Ph.D. Principals worked in all seven Israeli school districts, with 29 in elementary schools and 10 in middle schools.

Interview Measure and Procedure

Semi-structured interviews, where the interviewer developed and used an "interview guide" (i.e., list of questions and topics to be covered), were used for data collection. This enabled the researcher "to respond to the situation at hand, to the emerging worldview of the respondent, and to new ideas on the topic" (Merriam, 2009, p. 90). Key questions were preplanned, but the interviews were nevertheless conversational, with questions flowing from interviewees' previous responses whenever possible.

The interview aimed to explore principals' considerations during teacher evaluation. Thus, the interview included questions such as: Please try to remember a teacher you recently evaluated: How did you rate him/her? Why? How do you determine the ratings you give to an evaluated teacher? To what extent do grades you give in teacher evaluation reflect the teacher's true ability? With regard to the specific phenomenon under study – teachers' over-evaluation by principals – to promote study participants' authentic disclosure during the semi-structured interviews and to avoid the implication that interviewees had acted unprofessionally, these study participants were not explicitly asked whether they themselves had given teachers inflated ratings. However, toward the end of the interview, interviewees

who had not raised this issue either explicitly or implicitly during their interview discourse were then asked to address principals' possible inflation of teachers' ratings as a phenomenon that might describe Israeli principals in the recent years of the New Horizon teacher evaluations.

In line with ethical considerations, all participants were informed that their participation was voluntary, with full exit options at any point in time; no one opted out. They were assured of confidentiality (all names in this paper are pseudonyms.) and were asked to provide written consent, based on their understanding of the study's goal. Interviews generally lasted one hour and were audiotaped for later transcription and analysis.

Data Analysis

Data analysis consisted of a four-stage process – condensing, coding, categorizing, and theorizing. First, the necessary sorting and condensing were performed (Miles, Huberman, & Saldaña, 2014), seeking out the relevant utterances on the topic of teacher evaluation. At the second stage – coding – each segment of data (utterance) was coded according to its meaning (Tracy, 2013). A code was most often a word or short phrase that captures the primary content and essence of a portion of an interview transcript, ranging in magnitude from a single word to a full sentence to an entire page. Many of the same codes were used repeatedly throughout. This stage was both data-driven and theory-driven, as it was based on a-priori codes as well as on inductive ones, developed by direct examination of principals' views on teacher evaluation as they articulated them during their interviews (Flick, 2009; Marshall & Rossman, 2011; Rossman & Rallis, 2012). After having captured the essence of utterances in the second stage, in the third stage – categorizing – similar utterances, which shared some characteristic, were grouped together into "families." This organization enabled to arrive at their general meaning, thus defining and labeling categories accordingly. Finally, the theorizing stage aimed to reach an abstract conceptual construct of the categories derived in

the previous stage, and to decipher both how they were interconnected and how they influenced each other (Richards & Morse, 2013).

To ensure the appropriateness of sampling, data collection and data analysis, two educational leadership professors evaluated and critiqued the researcher's decisions, providing additional perspectives of their own regarding research design and data interpretation. In addition, the member check method (Koelsch, 2013) was used. All interviewees were asked to check both accuracy and interpretation. As in any qualitative exploration, attention was paid to how the researcher's background and personal experience might impact theoretical and methodological perceptions concerning the inquiry. As the importance of reflective journals in qualitative research has been recognized (Ortlipp, 2008), the researcher wrote a personal reflective research log throughout the study to ensure critical thinking.

Findings

With regard to the phenomenon under study – inflation of teachers' evaluations by principals, where almost all ratings are skewed to the highest end of the spectrum – 21 of the interviewees explicitly confirmed that they had over-evaluated their own teachers. For example: "Like most principals, I often give teachers grades that are higher than what they really deserved" (Lisa, an elementary school principal with seven years of experience); "Of course, my ratings are too high. There is no connection between teachers' rating and reality" (Noah, an elementary school principal with 13 years of experience); "When it comes to rating teachers, I am very generous. And anyone who knows how the education system works, should not be surprised" (Naomi, an elementary school principal with nine years of experience). An additional 10 interviewees implicitly revealed their over-evaluation tendency while explaining their considerations for evaluating teachers. Six more interviewees stated (toward the end of the interview) that this tendency is common among Israeli principals. Only

two interviewees said they always gave teachers exactly what they deserved, and they did not know what other principals were doing.

In line with the study's primary aim, qualitative analysis of the data deriving from the 39 current interviewees yielded four considerations taken into account by principals that might lead to higher evaluations than what principals think teachers really deserved: (1) time constraints/prioritization (low perceived value for high time investment); (2) evaluation's ineffectiveness for improving teaching (via teacher development or dismissal); (3) the imprecision of teacher evaluation measurements; and (4) impingement on interpersonal relationships. Although these four considerations were distinct, they were closely interrelated in principals' process of teacher evaluation.

Time Constraints and Prioritization

Interviewed principals' most frequent reason for giving higher ratings than what they thought teachers deserved was time constraints, mentioned by 17 participants. Interviewees perceived the teacher evaluation procedure required by the Ministry of Education to be a multi-step process demanding substantial management time. Inasmuch as they perceived themselves as lacking the intensive amounts of uninterrupted time required for this procedure, they performed the evaluation procedure only partially and, to be on the safe side, gave teacher higher ratings. Principals inflated ratings "to be fair," or "to avoid resentment of teachers". For example, Naomi, an elementary school principal with nine years of experience, was afraid of teachers' complaints. She said:

> The procedure of teacher evaluation takes about nine meetings. That is simply
>
> too much. I wish I had that time. I have no free time; nor do I have control over
>
> how I spend my time. So I do the procedure very partially, give the teacher a
>
> high rating so that she won't have any complaints, and thus save myself all that
>
> trouble.

However, the current qualitative data analysis revealed that for 12 out of the 17 principals who attributed their rating inflation to time constraints this reason did not stand alone but rather was linked inexorably with principals' perceptions about the ineffectiveness or even uselessness of the evaluation process (see next theme), which led to these principals' low prioritization of the full teacher evaluation process into their time schedule. Principals surely lack sufficient available time; yet, the current interviewees' attribution of rating inflation tendencies to time constraints seemed to stem from their belief that teacher evaluation failed to make the best use of principals' precious time. As implied by these principals, if they had considered teacher evaluation to make a significant contribution to their primary instructional leadership objective of improving teaching quality, they would have found time for it. David, an elementary school principal with nine years of experience, said: "I don't invest in this nonsense, which only takes my time and does not contribute anything to the school." Similarly, Alice, an elementary school principal with nine years of experience, asserted: "It takes a lot of time, which could be used for much more beneficial things."

These utterances illustrate how principals' low prioritization of completing a full and accurate teacher evaluation was determined simultaneously both by actual lack of time and by their perception that the needed time investment would be disproportionate to the process's potential for real, relevant effects on improving teaching quality, as presented next.

Evaluation's Ineffectiveness for Improving Teaching

The second consideration disclosed by the current interviewees as leading to teacher over-evaluation involved their perception of the required teacher evaluation process as ineffectual for improving actual instruction in the school, mentioned by 14 participants. These principals said that because evaluation is not conducive to actually improving teachers' work, they saw no need to provide accurate ratings and thus took the easier path and simply

gave teachers high ratings. For example, Patricia, an elementary school principal with 11

years of experience, claimed that the required evaluation was not an effective way to improve

the quality of teachers' work: "I give teachers high ratings because this procedure [of teacher

evaluation] focuses only on giving teachers a performance score. A teacher does not become

an expert as a result of being evaluated, so I save my time for improvement programs."

According to Patricia, inasmuch as the evaluation process was not accompanied by an

improvement plan that included support for teachers' professional development, there was no

point in investing effort into it.

Among the 14 principals who perceived teacher evaluation as an ineffective tool for

them as instructional leaders, two principals felt that evaluations could contribute nothing

new to veteran teachers. Four principals believed that evaluations could not make any impact

because "people do not change" (Margaret, an elementary school principal with 23 years of

experience) or "since they don't really want to make the change deep down, it will be very

hard to go the distance" (Lisa, an elementary school principal with seven years of

experience). Six principals pinpointed the ineffectuality of teacher evaluation due to Israel's

current teacher tenure policy, which restricts principals' ability to fire teachers. These

principals claimed that because tenured teacher dismissal is not feasible, the results of low

teacher evaluation can yield no significant usefulness. For example, Elizabeth, with eight

years of experience as an elementary school principal, said: "Tenure makes removal of poorly

performing teachers, who have actually been found in their evaluation to be ineffective in the

classroom, simply impossible. If evaluation is impotent, why should I give real scores?"

Donald, an elementary school principal with 12 years of experience, one of the two principals

mentioned above, who said they always gave teachers exactly what they deserved, also said:

"I rate teachers as I think they really deserve, because in my eyes this is my professionalism

as a principal. Does it help? To me it is too difficult to get rid of a bad teacher."

Importantly, three principals explicitly claimed that teacher evaluation is pointless because principals are not expected to deal with improving teaching quality, as argued by Noah, an elementary school principal with 13 years of experience: "I'm no longer a teacher, and I'm not in the teaching business now. The teachers know how to do their job, and my job is to make sure they have the conditions to succeed." Interestingly, these principals did not see improving the quality of instruction as a central component of their principalship role.

The Imprecision of Teacher Evaluation Measurements

The third reason described for the frequent inflation of teacher ratings was principals' perception that teacher evaluation methods do not yield a complete, accurate picture of teachers' actual quality and real-time functioning on the ground, as mentioned by 12 participants. Asserting that instruction is too complicated to be measured accurately, interviewees' utterances indicated that principals often raised ratings beyond what teachers deserved because principals believed that teacher evaluation cannot be performed in a way that precisely represents all relevant details about the teacher's quality.

Four interviewees claimed that reliable measurement is hindered by the existence of numerous, multiple paths to effective teaching. For example, after nine years of experience as an elementary school principal, Harry explained: "Great teachers do make a difference, but that doesn't mean they all do so in the same way. The evaluation system assumes that all the great teachers look the same, but this is not the case in reality." Likewise, Margaret, a veteran elementary school principal with 23 years of experience, argued that multiple characteristics and various qualifications are necessary to function as an effective teacher: "Good teachers are made up of a combination of hundreds of qualities. Each good teacher has her own unique mixture of these qualities."

In addition to the complexity of determining who is a good teacher, three interviewees believed that teachers' desired characteristics also depend on the specific educational context.

Pamela, an elementary school principal with seven years of experience as principal, said: "Exemplary teaching looks and sounds different across different classrooms. We should lift our eyes up from the list of indices and see whether classroom practice actually reflects the education we want for our students." Focusing on one example of contextual factors that should affect teacher evaluation, one middle school principal with 14 years of experience, George, discussed the differing characteristics necessary for rural versus urban schools: "I believe that schools in the countryside are substantially different. While residents of metropolitan areas are competitive and look mainly for results, our parents have other priorities. Therefore, we need different teachers."

In sum, these principals perceived the required teacher evaluation measurements to be ambiguous and generic rather than precise and context-specific. Considering the evaluation process to be untrustworthy, these principals gave teachers overly high ratings to be fair with their teachers and leave a margin of security.

Impingement on Interpersonal Relationships

Interviewees' final, fourth reason for often giving teachers high ratings was the desire to maintain positive relationships with teachers, mentioned by 11 participants. Principals feared that poor ratings could harm their good relations with teachers, especially because principals considered their good working and personal relationships to be crucial for school improvement efforts. Michael, an elementary school principal with 10 years of experience, articulated this view clearly: "My good relationship with teachers allows me to run the school successfully. I work hard to build such relationships and do not want teacher evaluation to spoil it." David, with six years of experience as a middle school principal, likened his school to a warm family atmosphere, which allowed no room for the conflict inherent to serious teacher criticism:

Our school is like one big family. Both students and teachers get a feeling that

they belong, feeling at home almost right away. People here feel respected and

cared for and if they don't, they'll tell us and we'll resolve it… If I were to make

a tough teacher evaluation, like I was asked to do, it might substantially change

our family-like community.

For such principals, teacher evaluation is not just pointless because of its ineffectuality

– it may even be considered an active hindrance to school work. Linda, a middle school

principal with 16 years of experience, gave voice to this argument:

Teachers should not be under a magnifying glass. Evaluation is stressful for the

teachers and therefore not only does it not improve their work but, to the

contrary, it harms it. I conduct teacher evaluation because I have to, but to

minimize the damage, I always give teachers high ratings to relieve the

pressure from this process.

Thus, interpersonal relationships were a priority for some of the principals not only

because of the importance attributed to principal-teacher relationships for school success, but

also because of principals' warm, long-term relationships with their teachers. In particular,

having a close affiliation with their teachers, principals did not want low ratings to prevent

those teachers' promotion to a higher salary level. Altogether, the current qualitative data

revealed that principals' over-evaluation of teachers took into account the risk of low ratings'

deleterious impact on highly valued principal-teacher relationships and on the less proficient

teachers themselves who, even when not functioning optimally, were nonetheless members of

the school "family."

Discussion

This study explored Israeli elementary and middle school principals' considerations

while evaluating their teachers, to elucidate why principals frequently give teachers ratings

that are higher than they think these teachers truly deserve. To recap, qualitative data analysis

yielded four considerations of principals who give their teachers overly high ratings – as related to principals' prioritization of time investment, evaluation's impotence to improve teaching, its imprecise measurement of teachers, and its negative effects on interpersonal relationships. These considerations led principals not only to the avoidance of low ratings, but also to inflation of high ratings.

The considerations found in this study suggested that overall, principals considered evaluation results as exaggerated, rating teachers higher than warranted, rather than accurate, inflated only to account for factors not captured in the evaluation instrument. Moreover, because of these considerations, principals often sought to cede instructional feedback as much as they could rather than provide such feedback through parallel systems external to the evaluation tool. These considerations are important for understanding principals' position as mid-level policymakers in determining the extent to which the intended teacher evaluation that is imposed by policies from above is actually implemented in the school, and for understanding how teacher ratings fit into principals' role as instructional leaders.

Principals' Four Reasons for "Buffering" Imposed Policy

Principals most strongly voiced the issue of time constraints (mentioned by 17 participants) as a reason for their over-evaluation of teachers, thereby substantiating one of the four considerations recently suggested by Kraft and Gilmour (2017) in their preliminary exploration – the intensive amounts of time needed to conduct evaluations and document teachers' performance as well as to write up improvement plans. However, the present principals' discourse added new perspectives to the complex influence of principals' time constraints on their involvement in teacher evaluation. The interviewees did openly describe their lack of sufficient uninterrupted time blocks to deal with the laborious, lengthy process of teacher evaluation and its consequences; however, these time considerations often appeared to be inexorably linked with the low value that principals attached to the teacher

evaluation process (due to its perceived ineffectuality, imprecision, or adverse effects), suggesting a low reward for the extremely high time investment needed. Put differently, time constraints were a major consideration because principals saw teacher evaluation as a low priority, saying it was "not very important," "not urgent," or "not the first thing to do." Thus, the reason of time constraints found in this study not only supports Kraft and Gilmour's (2017) findings but also broaden their discussion.

More broadly, this disproportion between time investment and perceived potential rewards for an externally imposed policy to supposedly promote school improvement may help elucidate why the widely campaigned push for school principals to act as instructional leaders has only partially become a reality in actual schools. Several researchers have claimed that principals lack sufficient time to engage directly in attempts to improve teaching and learning (Goldring et al., 2015; Murphy et al, 2016) largely because of ongoing structural limits on principals' time that pressure them to attend to other issues like building operations or student affairs (Camburn et al., 2010). Instructional leadership is acknowledged as requiring uninterrupted blocks of time, while principals' workday is characterized by fragmentation of activities and crisis-oriented issues (Murphy et al., 2016; Prytula et al., 2013). However, the current qualitative findings may help interpret these assertions by clarifying that, on the ground, principals' time schedule is prioritized according to activities' perceived actual value for improving the school's teaching and learning. Thus, although the New Horizon teacher evaluation process may have been presented to principals by high-level policymakers as a supposed instructional leadership behavior, it appears that these principals may not have viewed the imposed evaluation procedure as an authentic instructional leadership activity at all. As described next, the three additional considerations that emerged from this study about the disadvantages of the teacher evaluation tool and process, together, suggest that if principals had considered the New Horizon teacher ratings to be actual

significant contributors to teaching quality, they would have likely prioritized the steps needed to produce faithful, representative teacher ratings into their limited time.

The second reason that emerged in the current study for principals' over-evaluation of teachers – a perception of the teacher evaluation process as pointless because of its limited efficacy for improving teaching – may be seen as consistent with the position of some scholars, who claimed that alternative improvement strategies may yield more positive results and at a lower cost in terms of staff time and district funds compared to teacher evaluation (Fullan, 2011; Hallinger, Heck, & Murphy, 2014; Mehta & Fine, 2015). The current interviewees thought that the teacher rating process could not fulfill either of the two major aims of teacher evaluation, teacher development or teacher measurement (Marzano, 2012). The absence of an improvement planning component in the Israeli teacher evaluation system reinforces interviewees' claim that the system could not effectively bring about an amelioration in teachers' work. Moreover, regarding the teacher measurement goal (Marzano, 2012), the interviewees claimed that distinguishing between high and low performing teachers would be pointless because teacher tenure policy would prevent unsatisfactory tenured teachers' dismissal. As in many western countries, Israel's tenure policy restricts principals' ability to fire tenured teachers to the extent that most schools end up retaining their poorly performing teachers (Yariv & Coleman, 2005). To be noted, several principals argued that teacher evaluation is ineffective because improving teaching quality is not a top priority for school principals. It is striking that, despite the longstanding campaign advocating the instructional leadership approach (Shaked, 2018; Goldring et al., 2008, 2015; Murphy et al., 2016), these participants did not see improvement of instruction as a foremost concern for school leaders.

The third reason that interviewees cited for inflated teacher ratings – involving the perceived inadequacy of teacher evaluation methods to accurately capture actual teacher

functioning – may be seen as broadening Kraft and Gilmour's (2017) preliminary study findings that principals did not necessarily focus on evidence-based criteria for teacher evaluation, emphasizing teachers' motivation and potential, especially for novice teachers. The interviewed participants reported giving overly high ratings "to be on the safe side" due to the perceived imprecision (untrustworthiness) of teacher evaluation measurements. A full and realistic assessment of teachers' strengths and weaknesses requires principals' knowledge as to which capabilities are most important for effective teachers and which teaching methods are effective in which contexts. Thus, in order for instructional leaders to attend to more aspects of instruction and to engage more effectively in improving teaching and learning (Lochmiller & Acker-Hocevar, 2016; Steele, Johnson, Otten, Herbel-Eisenmann, & Carver, 2015), their knowledge base should incorporate content knowledge, pedagogical knowledge, curricular knowledge, and knowledge of learners (Spillane & Louis, 2002; Stein & Nelson, 2003).

The fourth deterrent to accurate teacher evaluation that emerged in this study – the perceived adverse effects of realistic teacher ratings on interpersonal relationships in the school – is similar to Kraft and Gilmour's (2017) finding, that principals are concerned about the discomfort and conflict that may arise in the aftermath of poor teacher ratings, especially in cases of teacher dismissal. These perceptions among the current interviewees uphold previous claims underscoring healthy principal-teacher relationships as constituting the basis for principals' effective instructional leadership (Robinson, 2010). The capacity to build good relationships with teachers may be seen as vital for instructional leadership, inasmuch as the influence of principals on students is mainly indirect (Murphy et al., 2016), through influence on teachers' practices that, in turn, influence student learning and results (Louis et al., 2010; Supovitz et al., 2010; Thoonen et al., 2012). Although teacher evaluation is not necessarily harmful to teacher-principal relationships (Marzano & Toth, 2013), 11 principals were

preoccupied with its potential damage.

Theoretical and Practical Implications

Overall, the conceptualization of school principals as local, mid-level policymakers serves as a relevant framework for understanding the current findings on principals' considerations with regard to teacher evaluation. The execution of teacher evaluation policies greatly depends on school principals (Levin & Datnow, 2012; Spillane & Kenney, 2012) because, in their role as local unofficial policymakers, principals actively fit the school's policy practices to their own perceptions (Diamond, 2012; Louis & Robinson, 2012). As mediating agents between the extra- and intra-school worlds (Kelchtermans et al., 2011; Maxcy et al., 2010; Salter, 2014), principals may not necessary completely accept ("bridge") or reject ("buffer") external national demands (Kohansal, 2015; Maxcy et al., 2010; Paredes Scribner, 2013; Rutledge et al., 2010). Instead, school principals may partially accept or modify them (Schechter & Shaked, 2017; Shaked & Schechter 2017; Spillane & Kenney, 2012).

In the current case, a substantial proportion of the interviewed principals appeared to implement the teacher evaluation policy only superficially, using inflated unrealistic ratings without internalization and real change, which would be regarded as buffering (Hall & Hord, 2015; Honig & Hatch, 2004). Thus, while the present study's school principals reported doing their duty to evaluate teachers, their buffering modus operandi prevented the teacher evaluation policy from achieving its goal of providing accurate data about individual teachers' instructional performance. When these school principals believed that teacher evaluation was ineffective, inaccurate, or even a hindrance to school work, they found ways not only to leave their "mark" on the policy received from above but even to make this policy meaningless.

Turning to practice, the current study's findings suggest that to establish a workable

teacher evaluation policy, principals' perceptions and concerns must be taken into account.

Teacher evaluation processes should be designed in accordance with current knowledge on

effective ways to evaluate teachers in different school contexts and should be complemented

with a requirement to create an evidence-based improvement program for teachers whose

results are not satisfactory (Marzano & Toth, 2013). Principals should be included at the

policymakers table when evaluations are being developed. The launch of a teacher

evaluation policy should be accompanied by support for principals to help them understand

and identify with the policy's goals, by listening to their reservations and their perceptions of

their own schools' idiosyncratic contexts, by providing the needed evidence-based

knowledge that principals may lack so that they can better understand the rationales for

teacher evaluation structures and procedures, and by discussing possible guidelines for

designing effective teacher improvement plans. Likewise, policymakers should leave room

for school principals' professional judgment. If the authorities put too much pressure on

schools to implement external policy, it will likely be ignored or excessively altered or

rejected. School superintendents should find a balance between ensuring that these priorities

are acted on and giving principals opportunities to exercise their discretion as mid-level

policymakers. Inasmuch as regulations seem to be a problem, a structured, standardized

teacher evaluation may be less effective than evaluation system that gives principals some

degree of autonomy. The current findings also may offer important implications with regard

to the intractability of existing tenure policies. Policymakers may adopt the OECD (2011)

recommendations to use more flexible terms of teacher employment and provide schools with

more autonomy in teacher personnel management. In all these steps, policies should not be

seen as only a top-down process; in fact, agents at all levels (e.g., district officials, principals,

middle leaders) should be involved in order to increase the likelihood of effective change

(Levin & Datnow, 2012).

Furthermore, this study's findings suggest that principals' perception of themselves as instructional leaders, who use teachers' evaluation to improve the quality of instruction, requires strengthening. Inasmuch as principals might render imposed policy meaningless, as demonstrated in this study, instructional leadership should not be seen as a mere list of tasks that the principal is obliged to perform. Hopefully, practical steps to attend to principals' concerns and provide the necessary knowledge and support may engender greater trust in policymakers' rationales and may enhance principals' increased acceptance of instructional leadership premises. Perhaps such trust- and knowledge-building steps will enable principals to accept that, to meet 21st century education goals, top priority should be given to improving teaching methods, while everything else is of lesser priority (Murphy et al., 2016; Walker & Slear, 2011). Specifically, principals should be explicitly exposed to the substantial body of evidence attesting to the role of teacher evaluation in upgrading teacher functioning, which should be emphasized as directly affecting student learning and achievement outcomes (Tuytens & Devos, 2017).

Considering the important implications for instructional leadership of principals' reluctance to rate their teachers according to the national policy demands, teacher evaluation could be a subject for explicit discussion in principal preparation programs. Aspiring principals should be assisted in acquiring practical evidence-based methods to effectively identify the characteristics of effective teachers. In fact, discussion of this issue may also be beneficial to inservice principals, as part of their professional development or mentorship programs. This may improve principals' will and ability to engage in both evaluation of and feedback to teachers.

Inasmuch as prior research on this topic has been scarce, the current study provides new data on principals' considerations in teacher evaluation; however, it has several limitations. First, future research to establish an intercoder reliability would do well to strengthen the

legitimacy of qualitative analysis. Second, this study explored only principals of elementary and middle schools, in which the Israeli teacher evaluation policy was implemented. Further research should complement this study by also exploring high school principals' considerations. Third, further research should explore if principals' considerations in teacher evaluation can be explained by participants' characteristics, such as gender, age, education, or school level (elementary/middle), as well as by teachers' characteristics (for example, principals might shy away from harder ratings when it comes to critiquing teachers who have more extensive experience or knowledge in a particular field). Fourth, since the data for this study were collected in a particular context, replicating this study in various socio-cultural contexts will enable generalization of the findings to broader populations, possibly substantiating their international validity. Additional longitudinal studies, including repeated interviews with the same school principals in order to explore how their considerations have evolved and unfolded throughout the evaluation policy implementation, would also be useful.

References

Alsobaie, M. F. (2015). The principal's relationship with teacher and development literacy of
elementary school students. *Journal of Education and Practice*, *6*(35), 149-155.

Blase, J., & Kirby, P. (2009). *Bringing out the best in teachers: What effective principals
do*. Thousand Oaks, CA: Corwin.

Camburn, E., Spillane, J., & Sebastian, J. (2010). Assessing the utility of a daily log for
measuring principal leadership practice. *Educational Administration Quarterly*, *46*(5),
707-737.

Capstones – The Israel Institute for School Leadership. (2008). *Perception of the principal's
role in the state of Israel: Report by the professional committee to formulate policy
recommendations for the Ministry of Education*. Jerusalem, Israel: Avney Rosha.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods
approaches* (4th ed.). Thousand Oaks, CA: Sage.

Diamond, J. B. (2012). Accountability policy, school organization, and classroom practice:
Partial recoupling and educational opportunity. *Education & Urban Society, 44*(2), 151-
182.

Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher
quality*. Washington, DC: Center for American Progress.

Donaldson, M. L., & Papay, J. P. (2015). Teacher evaluation for accountability and
development. In H. F. Ladd, & M. E. Goertz (Eds.), *Handbook of research in education
finance and policy* (pp. 174–193). New York, NY: Routledge.

Edgerson, D. E., Kritsonis, W. A., & Herrington, D. (2006). *The critical role of the teacher-
principal relationship in the improvement of student achievement in public schools of
the United States.* Retrieved from http://files.eric.ed.gov/fulltext/ED491985.pdf

Flessa, J. J. (2012). Principals as middle managers: school leadership during the

implementation of primary class size reduction policy in Ontario. *Leadership and Policy in Schools*, *11*(3), 325-343.

Flick, U. (2009). *An introduction to qualitative research* (4th ed.). Thousand Oaks, CA: Sage.

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform*. Melbourne, Australia: Centre for Strategic Education.

Goldring, E., Grissom, J. A., Neumerski, C. M., Murphy, J., Blissett, R., & Porter, A. (2015). *Making time for instructional leadership*. Retrieved from http://www.wallacefoundation.org/knowledge-center/Pages/Making-Time-for-Instructional-Leadership.aspx

Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: What influences principal practice? *Journal of Educational Administration*, *46*(3), 332-352.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444.

Hall, G. E., & Hord, S. M. (2015). *Implementing change: Patterns, principles, and potholes,* (4th ed.). New York, NY: Pearson.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement. An analysis of the evidence. *Educational Assessment, Evaluation, and Accountability*, *26*(1), 5-28.

Hallinger, P., & Wang, W. C. (2015). *Assessing instructional leadership with the Principal Instructional Management Rating Scale*. Dordrecht, Netherlands: Springer.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16-30.

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness.

American Journal of Education, *116*(4), 491-523.

Israeli Central Bureau of Statistics. (2013). *Statistical abstract of Israel* (No. 64). Jerusalem, Israel: Author.

Israeli Ministry of Education. (2015). *CEO guidelines: Evaluation procedure for teaching staff and administrators* (Hebrew). Retrieved from http://cms.education.gov.il/EducationCMS/Applications/Mankal/EtsMedorim/8/8-2/HoraotKeva/K-2014-3-2-8-2-23.htm

Kelchtermans, G., Piot, L., & Ballet, K. (2011). The lucid loneliness of the gatekeeper: exploring the emotional dimension in principals' work lives. *Oxford Review of Education*, *37*(1), 93-108.

Koelsch, L. E. (2013). Reconceptualizing the member check interview. *International Journal of Qualitative Methods*, *12*(1), 168-179.

Kohansal, R. (2015). Public school principals: Agents of bridging and buffering. *Journal of School Leadership*, *25*(4), 621-658.

Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly*, *50*(2), 279-304.

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting "the Widget Effect": Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, *46*(5), 234-249.

Le Fevre, D. M., & Robinson, V. M. J. (2015). The interpersonal challenges of instructional leadership: Principals' effectiveness in conversations about performance issues. *Educational Administration Quarterly*, *51*(1), 58-95.

Levin, J. A., & Datnow, A. (2012). The principal role in data-driven decision making: Using case-study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement*, *23*(2), 179-201.

Lochmiller, C. R., & Acker-Hocevar, M. (2016). Making sense of principal leadership in

content areas: The case of secondary math and science instruction. *Leadership and

Policy in Schools*, *15*(3), 273-296.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student

achievement? Results from a national US survey. *School Effectiveness and School

Improvement*, *21*(3), 315-336.

Louis, K. S., & Robinson, V. M. (2012). External mandates and instructional leadership:

school leaders as mediating agents. *Journal of Educational Administration*, *50*(5), 629-

665.

Marchant, G. J., David, K. A., Rodgers, D., & German, R. L. (2015). State teacher evaluation

and teacher education. *Teacher Educator*, *50*(2), 89-108.

Marshall, C., & Rossman G. B. (2011). *Designing qualitative research* (5th ed.). Thousand

Oaks, CA: Sage.

Marzano, R. J. (2012). Teacher evaluation: What's fair? What's effective? *Educational

Leadership*, *70*(3), 14-19.

Marzano, R. J., & Toth, M. D. (2013). *Teacher evaluation that makes a difference: A new

model for teacher growth and student achievement*. Alexandria, VA: ASCD.

Maxcy, B., Sungtong, E., & Nguyen, T. (2010). Challenging school leadership in Thailand's

southern border provinces. *Educational Management, Administration & Leadership*,

38(2), 164-183.

May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction.

Educational Administration Quarterly, *47*(2), 332-352.

Mehta, J., & Fine, S. (2015). Bringing values back in: How purposes shape practices in

coherent school designs. *Journal of Educational Change*, *16*(4), 483–510.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San

Francisco, CA: Jossey-Bass.

Miles, M. B., Huberman, M. A., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Murphy, J., Neumerski, C. M., Goldring, E., Grissom, J., & Porter, A. (2016). Bottling fog? The quest for instructional management. *Cambridge Journal of Education*, *46*(4), 455-471.

OECD - The Organisation for Economic Co-operation and Development. (2009). *Teacher evaluation: A conceptual framework and examples of country practices*. Retrieved from http://www.oecd.org/edu/school/44568106.pdf

OECD - The Organisation for Economic Co-operation and Development. (2011). *Teachers matter: Attracting, developing and retaining effective teachers*. Retrieved from http://www.oecd.org/edu/school/48627229.pdf

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, *13*(4), 695-705.

Paredes Scribner, S.M. (2013). Beyond bridging and buffering: Cases of leadership perspective and practice at the nexus of school–community relations. *Journal of Cases in Educational Leadership*, *16*(3), 3-6.

Price, H. E. (2015). Principals' social interactions with teachers: How principal-teacher social relations correlate with teachers' perceptions of student engagement. *Journal of Educational Administration*, *53*(1), 116-139.

Price, H. E., & Moolenaar, N. M. (2015). Principal-teacher relationships: Foregrounding the international importance of principals' social relationships for school learning climates. Journal of Educational Administration, 53(1), 116-139.

Prytula, M., Noonan, B., & Hellsten, L. (2013). Toward instructional leadership: Principals perceptions of large-scale assessment in schools. *Canadian Journal of Educational*

Administration and Policy, *140*, 1-30.

Richards, L., & Morse, J. M. (2013). *Readme first for a user's guide to qualitative methods* (3rd ed.). Thousand Oaks, CA: Sage.

Robinson, V. M. J. (2010). From instructional leadership to leadership capabilities: Empirical findings and methodological challenges. *Leadership and Policy in Schools*, *9*(1), 1-26.

Rossman, G. B., & Rallis, S. F. (2012). *Learning in the field: An introduction to qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Rutledge, S. A. Harris, D. N., & Ingle, W. K. (2010). How principals "bridge and buffer" the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring. *American Journal of Education*, *116*(2), 211-242.

Salter, P. (2014). Knowing Asia: Creative policy translation in an Australian school setting. *Journal of Education Policy*, *29*(2), 145-164.

Schechter, C., & Shaked, H. (2017). Leaving fingerprints: Principals' considerations while implementing education reforms. *Journal of Educational Administration*, *55*(3), 242-260.

Shaked, H. (2018). Why principals sidestep instructional leadership: The disregarded question of schools' primary objective. *Journal of School Leadership*, *28*(4), 571-538.

Shaked, H., & Schechter, C. (2017). School principals as mediating agents in education reforms. *School Leadership & Management*, *37*(1-2), 19-37.

Spillane, J., & Kenney, A. W. (2012). School administration in a changing education sector: The U.S experience. *Journal of Educational Administration, 50*(5), 541-561.

Spillane, J. P., & Louis, K. S. (2002). School improvement process and practices: Professional learning for building instructional capacity. In J. Murphy (Ed.), *The educational leadership challenge: Redefining leadership for the 21st century* (pp. 83-104). Chicago, IL: University of Chicago.

Steele, M. D., Johnson, K. R., Otten, S., Herbel-Eisenmann, B. A., & Carver, C. L. (2015). Improving instructional leadership through the development of leadership content

knowledge: The case of principal learning in algebra. *Journal of Research on Leadership Education*, *10*(2), 127-150.

Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis*, *25*(4), 423-448.

Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post NCLB era. *Education Finance and Policy*, *11*(3), 340–359.

Stronge, J. H., Richard, H. B., & Catano, N. (2008). *Qualities of effective principals*. Alexandria, VA: Association for Supervision and Curriculum Development.

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, *46*(1), 31-56.

Taylor, S. L., Bogdan, R., & DeVault, M. L. (2016). *Introduction to qualitative research methods: A guidebook and resource* (4th ed). Hoboken, NJ: John Wiley & Sons.

The New Teacher Project (2007). *Hiring, assignment, and transfer in Chicago public schools*. New York: The New Teacher Project.

Thoonen, E. E., Sleegers, P. J., Oort, F. J., & Peetsma, T. T. (2012). Building school-wide capacity for improvement: The role of leadership, school organizational conditions, and teacher factors. *School Effectiveness and School Improvement, 23*(4), 441-460.

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Chichester, UK: Wiley-Blackwell.

Tschannen-Moran, M., & Gareis, C. R. (2015). Faculty trust in the principal: An essential ingredient in high-performing schools. *Journal of Educational Administration*, *53*(1), 66-92.

Tuytens, M., & Devos, G. (2017). The role of feedback from the school leader during teacher

evaluation for teacher and school improvement. *Teachers and Teaching: Theory and*

Practice, *23*(1), 6-24.

Walker, J., & Slear, S. (2011). The impact of principal leadership behaviors on the efficacy of

new and experienced middle school teachers. *National Association of Secondary School*

Principals Bulletin, *95*(1), 46-64.

Wechsler, M., Tiffany-Morales, J., Campbell, A., Humphrey, D., Kim, D., Shields, P., &

Wang, H. (2007). *The status of the teaching profession 2007*. Santa Cruz, CA: Center

for the Future of Teaching and Learning.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national*

failure to acknowledge and act on differences in teacher effectiveness. Washington,

DC: New Teacher Project.

Yariv, E., & Coleman, M. (2005). Managing "challenging" teachers. *International Journal of*

Educational Management, *19*(4), 330-346.