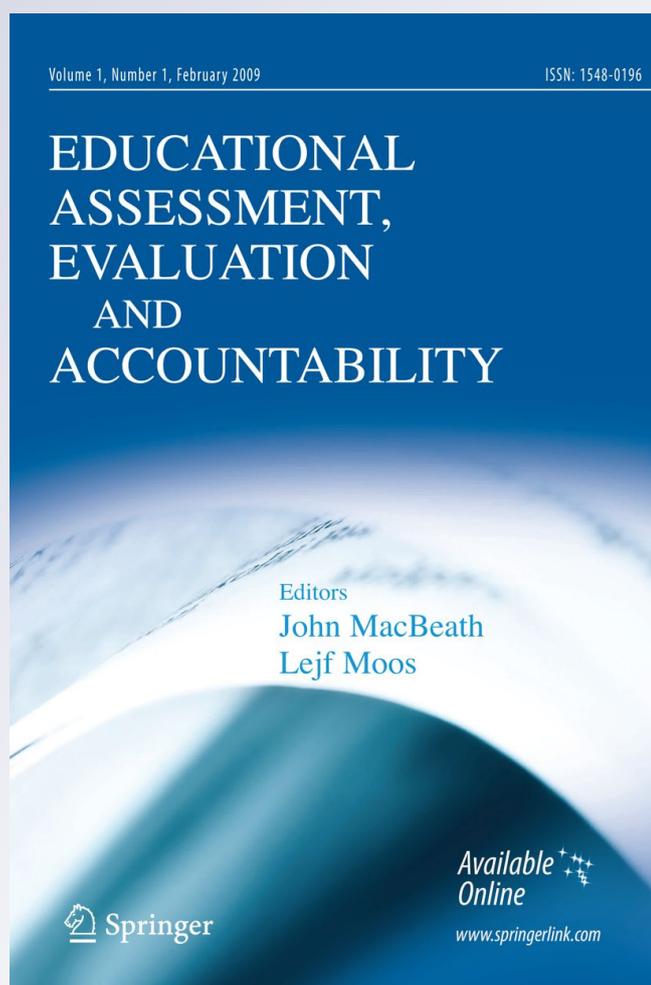*Association of District Principal Evaluation with learning-centered leadership practice: Evidence from Michigan and Beijing*

Min Sun, Peter Youngs, Haiyan Yang, Hongqi Chu & Qian Zhao

Springer

Association of District Principal Evaluation with learning-centered leadership practice: Evidence from Michigan and Beijing

Min Sun · Peter Youngs · Haiyan Yang · Hongqi Chu · Qian Zhao

Abstract Principal evaluation has become a key component of national policy debates on developing effective leaders. To contribute to these debates, this study draws on survey data to explore how principals in Michigan and metropolitan Beijing behaved differently in enacting leadership related to teaching and learning, and how they were evaluated differently by districts in these two regions in 2007–08 (Michigan) and 2008–09 (Beijing). We further combine these two independent samples and build two-level Multivariate Hierarchical Linear Models (HMLM) to estimate the extent to which district evaluation features were associated with principal leadership practice. Descriptive results indicated that Beijing principals were more likely to engage in core activities of instructional leadership (such as direct interactions with students about their learning, observing classrooms, and providing feedback to teachers) than their Michigan peers. Moreover, HMLM results suggest that district principal evaluation can serve as a powerful policy instrument to promote instructional leadership and should emphasize principals' organizational impact on instructional and learning outcomes in determining evaluation purposes, contents, and sources of evidence.

Keywords Principal leadership · District principal evaluation · Developing effective leaders · Empirical study

M. Sun (✉)
Department of Educational Leadership and Policy Studies, Virginia Tech, Room 206 East Eggleston, Blacksburg, VA 24061, USA
e-mail: sunmin@vt.edu

P. Youngs
Department of Teacher Education, Michigan State University, East Lansing, MI 48824, USA

H. Yang
College of Education, Renmin University of China, Beijing, China

H. Chu · Q. Zhao
Faculty of Education, Beijing Normal University, Beijing, China

In the United States, recent accountability policies in K-12 education have placed greater responsibility on principals for initiating reforms and enhancing student achievement (Bryk et al. 2010; Heck and Marcoulides 1996; Leithwood et al. 2004; Mangin 2007; Marks and Nance 2007; Stein and Spillane 2005). As a result, there have been many debates about how to hold principals accountable for student learning through principal evaluation and how to support and develop principals' ability to promote instructional improvement centered on student learning. Although researchers have presented theoretical arguments concerning the purposes, procedures, and instruments for evaluating principals, there is little empirical evidence of relationships between these features and actual principal leadership activities (Sun and Youngs 2009; Portin et al. 2006). Moreover, although researchers have recently developed new evaluation instruments (e.g., the Vanderbilt Assessment for Leadership in Education (VAL-ED™)) and more districts have tied student achievement to principal performance, little is known about the impact of district principal evaluation (e.g., purposes, content, timing, and consequences of evaluation) on principal leadership behavior (Sun and Youngs 2009).

In addition, few studies have compared principal leadership behaviors between the United States (U.S.) and mainland China (Cravens 2008) and no prior studies have compared the features of district principal evaluation practices between these two countries. The results of the 2009 Organization for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) indicate that Chinese students in elementary and secondary schools have significantly higher performance in reading, mathematics, and sciences than international peers who took the same PISA test, including peers from the U.S. (OECD 2010). The OECD report further suggests that school governance can affect students' reading performance (OECD 2010). China is one of the countries where there has been a long tradition of posting achievement data publicly and having system-level accountability for students' learning. Therefore, it is useful to probe into one aspect of the school governance and accountability system, school leadership, to examine a) how Chinese principals enact instructional leadership, and b) how principals are held accountable for student learning outcomes through district personnel evaluation.

By drawing on survey data from 88 K-12 principals in Michigan (the U.S.) and 90 K-12 principals from metropolitan Beijing (China), this study explores how principals behaved differently in enacting leadership related to instruction and learning, and how they were evaluated differently by districts in these two regions in 2007–08 (Michigan) and 2008–09 (Beijing). The sample from Beijing, to a large extent, represents general practices of principal leadership and district evaluation due to the centralized and uniform school governance system in China. However, the Michigan sample may not well represent principal leadership activities and district evaluation across states in the U.S. because of the long tradition of local control of educational practices by individual states. Since limited literature that focuses solely on either Michigan or Beijing district principal evaluation and leadership practices, we extend our literature review to these two countries, the U.S., and China, to situate our study into broader contexts.

Next, we combine these two independent samples from Michigan and Beijing, which greatly increases the variance of principal leadership practices and that of district evaluation practices. We build two-level Multivariate Hierarchical Linear

Models (HMLM) to estimate the extent to which district evaluation features were related to principal leadership behavior. The results provide new empirical evidence concerning how to develop effective school leaders to support teaching and learning.

1 Principal leadership in teaching and learning

Effective leadership improves teaching and learning (Leithwood et al. 2004). While some evidence about principal leadership effects on student learning can be difficult to interpret (e.g., Hallinger et al. 1996; Hallinger 2011), recent studies by North American scholars showed that the total (both direct and indirect) effects of leadership on student learning accounted for about one fifth of the total variation (20%) explained by all school-level variables (Creemers and Reezigt 1996), after controlling for student personal and family background factors. Principal leadership has been reported as the second most important factor influencing student learning within schools, following teachers' impact (Leithwood et al. 2004). The evidence supported the widespread interest in the last two decades in identifying essential ingredients of successful leadership in terms of improving teaching and learning. For example, Hallinger and Heck (1999) described effective leadership practices as addressing "purposes," "people," and "structures and social systems." Leithwood's (1996) categories of successful leadership practices included "setting directions," "developing people," and "redesigning the organization." Murphy (2007) argued that effective leadership behaviors were those that focused on student learning. Although the labels may slightly differ, the essence of effective leadership identified in the literature includes the practices elaborated as follows.

Leithwood and Janzi (2005)'s study documented evidence of how leaders should establish rigorous student learning goals that were achievable and compelling. School leaders can strengthen instructional quality and student learning by developing clear visions and goals, which can be translated into grade-level objectives (Newmann et al. 2000). School leaders can build support for these ambitious goals among school staff and articulate each teacher's responsibility for helping students to meet them (Hallinger and Heck 1996; Leithwood 1994). They can also coordinate curriculum in ways that ensure all students have access to rigorous academic content in language arts, mathematics, and other subject areas by helping teachers translate the school's academic goals into common curricular objectives and instructional activities (Murphy et al. 2007a, b; Newmann 1997).

With access to student achievement data, principals should be able to effectively use such data to diagnose problems and work with teachers to implement actions to promote instructional changes (Bottoms 2001; Heck 1992; Leithwood and Janzi 2005). Beyond distributing information about students' achievement on state assessments, principals should encourage teachers to develop formative assessments and performance tracking systems within their schools. Principals can oversee the quality of instruction, which is defined as teaching practices that maximize student engagement and achievement, through classroom observations, post-observation conferences, and other direct contact with teachers (Burch and Spillane 2003; Coburn 2001; Stein and D'Amico 2002). Moreover, leaders can recognize students who do superior

academic work with formal and informal rewards to motivate them to strive for academic excellence (Heck et al. 1990).

Effective leaders initiate strategies to develop teachers in schools. Beyond direct on-site coaching, leaders can monitor programs for junior faculty members, send teachers to off-campus professional development programs, and distribute journals to help teachers update their knowledge and skills (e.g., Payne and Wolfson 2000; Stein and Nelson 2003; Walker and Slear 2011). In addition, effective principal leadership includes promoting instructional coherence (Newmann et al. 2001) and fostering productive work relationships among teachers and between themselves and teachers (Sun et al. 2010; Sun et al. 2011). In particular, such leadership involves promoting teacher professional community (Louis et al. 1996) and relational trust between teachers and school administrators and between school staff and students and their families (Bryk and Schneider 2002). Moreover, acknowledging teachers' effective practices formally and informally is another important strategy for motivating teachers (Hipp 1996; Marzano et al. 2005).

All of the above components of effective principal leadership have been echoed by scholarship in China, where there is a widespread expectation that principals should be strong instructional leaders (e.g., Chen 2011; Chu and Liu 2010; Xu 2010; Zhao and Liu 2010). Similar to principal leadership in North American contexts, instructional leadership in China also involves setting standards for student learning, optimizing teaching environment, directing and evaluating teaching, and promoting teachers' professional growth (Cravens 2008). After reviewing Chinese scholarship on instructional leadership, we find that many of these discussions are centered on leadership activities within school buildings and even within classrooms. For example, Chu and Liu (2010) discussed the ways in which and how often principals should teach regular classes, visit classrooms, and evaluate instruction. Chen (2011) administered surveys to 745 teachers in eight provinces and found that teachers perceived that their principals placed greatest priority on a) encouraging teachers to conduct research and reflect on their instruction, and publish these research results and self-reflections; b) encouraging teachers to observe each other's instruction and learn from each other; and c) visiting classrooms, evaluating instruction, and providing feedback on teachers' instruction. He also found group differences in these behaviors. Female principals more frequently engaged in these behaviors than male principals and principals who had a Bachelor's degree or higher more often enacted these leadership behaviors than other peers. However, there were no statistically significant differences among principals who came from various geographical locations (e.g., Western, Eastern, or Middle China) or based on urbanicity (e.g., urban, suburban, or rural settings).

Research studies that use the same instruments to compare the leadership practices of U.S. and Chinese principals are scarce. Jing (2010) administered the same survey instrument to two independent samples of U.S. (18 schools in Phoenix, AZ) and Chinese principals (18 schools from Harbin, China) and found some differences in instructional leadership activities between these two groups. In both samples, school administrators reported that they had engaged more frequently in "developing people" and "building trust" than other leadership activities. But in American sample, principals rated "visioning" and "sharing power and influence" as their most frequent practices while assistant principals most frequently engaged in developing people and trust-building. In contrast, Chinese principals shared such responsibility with assistant

principals (Jing 2010). Although this study explicates some of the similarities and differences between American and Chinese high school principals and assistant principals, the instrument used in the study was framed under the concept of "shared leadership" rather than "learning-centered leadership" and the sample was limited to high schools. The field still lacks an empirical description of the ways in which Chinese principals enact instructional leadership that promotes the strong performance of Chinese students on international tests. Probing into the specific practices of Chinese principals, in particular, and comparing them to those of their peers in the U.S. have the potential to reveal ways by which instructional leadership in both countries can contribute to improved student learning. Moreover, our data collected in this study allow us to investigate factors that can explain variation in principals' instructional leadership practices.

2 District evaluation that supports the development of effective principal leadership

District-level evaluation of principals is one of the common strategies used to hold principals accountable for school performance, to identify principals' professional development needs, and to track principals' progress over time. District personnel evaluation has a long history in both the U.S. and China in terms of using relatively simple devices, such as an annual visit by a district administrator that might feature a checklist of behaviors (Stine 2001). We review the features of these district evaluation practices in terms of their purposes, contents, and procedures.

2.1 Purposes

In the United States, the practice of principal evaluation is common (Doud and Keller 1998) and has been used to make decisions about hiring, promotion, (re)assignment, and contract renewal; that is, to satisfy personnel policy requirements. Moreover, principal evaluation can be used as a diagnostic tool to identify an employee's strengths or areas in need of improvement (Stine 2001). These formative purposes of evaluation have been widely emphasized in many local districts. At the same time, being an integral part of current standard-based accountability systems, the summative purposes of principal evaluation hold principals accountable for student achievement and develop their capacity to facilitate school restructuring.

In China, school principals have long been evaluated and compensated as governmental officials or managers of teachers. The official ranking of a principal is determined by the status of the school the principal worked for and their rank, in turn, determines his/her income (Chu and Yang 2009; Qian 2009). Leadership evaluation is often regarded as one aspect of school evaluation (Guo 2005). The main purposes of principal evaluation include a) monitoring principals' performance and holding them accountable for school performance; b) providing evidence for determining merit salary increases (Xin and Wang 2006) or reassigning leaders to schools with higher or lower rankings; and c) serving as a formative assessment to identify principals for off-campus professional development or on-campus mentoring or coaching (Guo 2005).

2.2 Contents

After analyzing 94 district evaluation instruments used in the United States, Goldring and her colleagues (2008) found that there are generally four approaches that have been used in the U.S.: a) evaluating the extent to which principals have accomplished the functions, tasks or responsibilities of the position, such as managing school programs, pupil personnel, community relations, physical facilities, and student behavior, and coordinating professional development; b) assessing key competencies, knowledge and skills that principals should possess, such as knowledge related to instruction and management, dispositions, and leadership styles; c) evaluating leadership processes and practices covering effective leadership domains, such as the *Interstate School Leaders Licensure Consortium* (ISLLC) standards; and d) assessing outputs, such as student achievement and other organizational outcomes related to school improvement. With the development of accountability systems, the last approach has become increasingly popular, but researchers have warned about the possible risks of relying solely on outcome-based assessments, but ignoring organizational and contextual factors that can help explain student achievement and other organizational outcomes (Goldring et al. 2008).

In China, the four aspects of school leadership recommended by the Ministry of Education in 1992 are "values" (r.f. moral and ethical leadership), "abilities" (r.f. management skills and knowledge), "diligence" (r.f. diligent leadership behaviors), and "achievement" (r.f. leadership performance as measured by student achievement). All four of these aspects have been enacted at the local district level with a strong emphasis on "values" and "achievement". Highly influenced by Confucian philosophy, principals themselves and their supervisors believe that being a moral leader is a key to successful leadership (Qian 2009). Moreover, there is widespread belief that leadership can impact student performance on municipal or provincial common final exams as well as graduation exams and rates (Cravens 2008). In the mid-2000s with the development of the concept of principal professionalism (e.g., Chu 2003; Chu and Yang 2002) and major reform efforts in China that addressed a number of issues including social learning, creative thinking, and physical health, principal evaluation systems now emphasize the strategic development of teachers and students, achievement, and social relations. Further, instructional leadership has become the core of new principal evaluation systems (Chu and Cravens 2011).

Moreover, Chinese researchers have pointed out that it is unfair to base principal evaluation entirely on evaluations of school performance because there are many factors affecting student achievement beyond principals' control, such as parental involvement in students' education after school (e.g., Xin and Wang 2006). Moreover, with the growing belief that principals should be recognized as professionals rather than government officials (e.g., Xin and Wang 2006), districts in Shanghai have implemented career ladder system to evaluate and compensate principals. Such systems rank principals based on their working experiences; leadership knowledge, skills, and practices; and organizational outcomes (Yang 2006).

2.3 Procedures

In the U.S., while the district superintendent is usually responsible for evaluation, this duty is often assigned to other district administrators, such as the assistant

superintendent of instruction, the director of curriculum, or the direct of personnel management. In addition, there is a growing trend of involving parents, teachers, and principals themselves. Principals normally are assessed once a year, but sometimes every other year or every 3 years. Further, portfolios and/or open, narrative-form evaluations are often used to help document principals' performance against a set of predetermined goals (Doud and Keller 1998; Lashway 2003; Stine 2001). One study found that the evaluation formats and processes frequently differed from one district to another; and that most principals believed that their evaluations were not useful and were influenced by political forces beyond their control (Davis and Hensley 1999). Reeves (2005) drew on a survey of principals throughout the U.S. and reported that few school leaders felt that their motivation or performance was shaped by principal evaluation. Most principals in this study believed that the criteria used to evaluate them were not well-specified, that there were few consequences attached to the evaluation process, and that it did not provide helpful feedback to them.

In China, principals are appointed and evaluated by district administrators at the beginning of every school year. They use similar procedures and tools as the U.S. district administrators, such as written documentation of principals' actions and behaviors, observations during faculty and parent meetings, principals' oral presentations, and teachers' ratings of principals' performance (Xin and Wang 2006). Regardless of whether principals are evaluated as government officials or under the new career ladder schemes, results from leadership evaluation can have serious implications for principals' career development, income, and reputation. However, similar to criticisms of American principal evaluation, the evaluation criteria are usually not specific. The evaluators often lack school management and teaching experiences and thus can't always accurately evaluate principals' performance, let alone give them useful feedback (e.g., Guo 2005; Xin and Wang 2006).

Most of studies summarized above have been conducted by separate researchers in each of these two countries. None of the prior studies has systematically collected comparable data on the features of district principal evaluation practices. Through this thorough literature review, we find many similarities in purposes and content foci of district evaluation, but differences in associated consequences for principals. To better articulate these similarities and differences, this study collected empirical data from random samples of principals in Michigan and Beijing metropolitan area by using the same survey instrument.

2.4 The association between features of district principal evaluation
and leadership practices

In the U.S., principal evaluation has been criticized for being poorly aligned with educational priorities, such as improving student achievement, accountability, and school change (e.g., Linn 2000; Marcoulides et al. 1995). Evaluation policies and practices are often based on little empirical evidence that such evaluation practice would promote effective leadership centered on teaching and learning (Goldring et al. 2007). Recognizing these limitations in practice, few research efforts have been undertaken to address the development of principal evaluation practices that focus on the outcome of student learning and link features of evaluation with leadership behaviors that have the potential to promote student learning.

Based on the conceptual model of learning-centered leadership, Murphy, Goldring, and Porter led a group of researchers to develop the VAL-ED™. This new instrument is designed to validly and reliably measure six core components and six key processes related to learning-centered leadership behavior (refer to Murphy et al. Murphy et al. 2007a, b, for details). Although the contents of VAL-ED™ have been supported by solid existing scholarship and practice, few studies have verified the effect of the use of this instrument on principals' learning-centered leadership behaviors. Furthermore, while the Murphy et al.'s study focused on the development of the content and psychometric properties of the assessment instrument, other features of district evaluation policy, such as purposes, timing, consequences, were not explored in their project.

In a separate effort, Kimball et al. (2007) carried out a randomized design study of principal evaluation and school leader outcomes in a large district in the western U.S. For the study, 44 principals were randomly assigned to a new standards-based performance evaluation system and 44 principals were evaluated under the district's old system.[1] While the researchers reported several important outcomes that seemed to be associated with use of the new evaluation system, few of their results were statistically significant due to low response rates and the district's inability to fully enact the random assignment. The study outcomes included more clarity for principals about the expectations related to evaluation and more detailed feedback as part of the evaluation process. In addition, those assigned to the new system seemed to place more priority on technology and communicating a school vision than those assessed under the old system. Finally, principals evaluated under the new system seemed to receive more assistance with improving their leadership practices (Kimball et al. 2007). This study examines the impact of the whole evaluation system, not individual features. Furthermore, the small sample sizes make the results less likely to be generalized to a larger population with more diverse contexts.

In a third study, Sun and Youngs (2009) employed Multivariate Hierarchical Linear Models (HMLM) to examine relationships between principals' behaviors and district principal evaluation purpose, focus, and assessed leadership activities in 13 Michigan districts. The study reported that principals were more likely to engage in learning-centered leadership behaviors when the purposes of evaluation included principal professional development, school restructuring, and accountability; when the focus of evaluation was related to instructional leadership; and when evaluation addressed leadership in school goal setting, curriculum design, teacher professional development and evaluation, and monitoring student learning.

Our prior study (Sun and Youngs 2009) only collected data from principals in Michigan. To increase the variation and diversity in the previous sample, we add a new sample of principals from the Beijing metropolitan area and include more measures

---

[1] The district's new principal evaluation system was based on its teacher evaluation system, which, in turn, was based on Danielson's Framework for Teaching (1996); the new system was characterized by clear performance standards (dimensions) and rubrics differentiating performance on the dimensions. Depending on a given principal's years of experience and prior evaluations, he or she was evaluated with regard to their performance on two or more of the dimensions each year. Under the new evaluation system, most of the data were collected by district administrators through observations and conferences. The district's old evaluation system, like those in many other districts, consisted of a checklist of personal traits and behaviors, and it did not feature a rubric differentiating levels of performance (Kimball, Milanowski, and McKinney 2007).

of evaluation features and principals' practices. Although we descriptively compare evaluation practices and leadership behaviors in these two regions, the central purpose of this study is to empirically explore the features of principal evaluation that have the potential to lead to desired leadership behaviors that focus on students' learning processes and outcomes.

3 Methods

3.1 Data sources

We collected data on district principal evaluation policy and principal leadership practices in Michigan (the U.S.) and the Beijing metropolitan area (China). In spring 2008, we administered surveys to a random sample comprising138 K-12 public school principals in Michigan who had served in 2006–07 as principals in their buildings and were continuing to serve in this role in 2007–08. In fall of 2009, we administered the same surveys to a random sample of 145 K-12 public school principals in metropolitan Beijing (see the next page for more information about the survey development).[2] We received valid responses from 88 Michigan principals and 90 principals in metropolitan Beijing. Therefore, our response rates were 63.8% (Michigan) and 62.1% (Beijing), respectively. Because there are differences in financial conditions and parental involvement between Beijing city schools and those located just outside of Beijing, we accounted for this in our data analysis process.

On average, these two samples were comparable with regard to demographic characteristics, as reported in Table 1. About 57.9% of Michigan principals worked in elementary schools, 40.9% worked in secondary school settings, and 1.2% worked in other K-12 settings (such as Kindergarten etc.). Comparably, in Beijing, 53.3% of the principals worked in elementary schools, 30% worked in middle or high schools, and the rest worked in other K-12 settings. Slightly less than half of each sample was female and less than half had worked as a principal for more than 3 years.

However, these two samples differed with regard to the percentage that had more than 3 years of work experience as full-time teachers (p-value of the *t*-test ≤0.0001). About 62% of principals in Michigan had more than 3 years of working experience as full-time teachers while 95% of principals in Beijing had such teaching experience.

We administered the same survey to Michigan principals (in English) and Beijing principals (in Chinese). We first developed the English version of the instrument through in-depth interviews with two Michigan K-12 principals and through two rounds of pilot studies undertaken during the 2007–08 school year. We then translated this English version into Chinese. The first author has had extensive experience working with principals in both countries and took the lead role in translation to make sure that these two instruments were equivalent and the Chinese version was adaptable to Chinese school contexts. The second author's professional experience in the U.S. and the last three authors' professional experience in China greatly enhanced the instrument's cross-context equivalence and within-context appropriateness. We then invited 10 principals

---

[2] Since there was no major educational reform newly introduced between 2008 and 2009 in either Michigan or Beijing, the one-year difference in collecting data will not affect this comparison.

Table 1 Basic Sample Characteristics

	MI	BJ
Percentage of elementary school principals	57.95	53.33
Percentage of middle or high school principals	40.91	30
Percentage of female principals	44.32	42.31
Percentage of principals who held a Bachelor's degree or higher	100	91.46
Percentage of principals who had more than 3 years of working experience as principals	49.38	42.31
Percentage of principals who had more than 3 years of working experience as full-time instructional teachers (a)	62.2	95

MI=Michigan, BJ=Beijing

(a) There was a significantly higher percentage of Beijing principals who have had more than 3 years of working experiences as full-time instructional teachers ($p \leq 0.0001$) than that of Michigan principals

in a preparation program at Beijing Normal University to review the Chinese version of the instrument and we subsequently revised it based on their feedback.

The final version of the principal survey had three sections. The first section inquired into participants' perceptions of features of district's principal evaluation system; the second section focused on learning-centered leadership behaviors; and the third section asked about school contexts and principals' personal and professional backgrounds.

3.2 Measures

We aggregated composite measures of a) principals' leadership practices and b) features of district principal evaluation based on factor analysis, as indicated respectively in Tables 2 and 3. We used internal consistency ratings as indicators of the reliability of these composite measures for the whole sample and two subsamples. The Cronbach's α for the whole sample is indicated as Cronbach's α_{Whole}, the Cronbach's α for the Michigan sample is indicated as Cronbach's α_{MI}, and the Cronbach's α for the Beijing sample is indicated as Cronbach's α_{BJ}.

Table 2 reports composite measures and associated survey items as well as the internal consistency of reliability for principal learning-centered leadership for the whole sample and two subsamples. These composite measures include "framing school goals," "communicating school goals," "supervising and evaluating instruction," "coordinating the curriculum," "monitoring students' progress," "protecting instructional time," "maintaining high visibility," "providing incentives for teachers," and "providing incentives for student learning." The Cronbach's αs for these composite measures are consistently higher than 0.7 for the whole sample, although some of the measures are less than 0.6 for the subsamples. For instance, the composite measure of "supervising and evaluating instruction" for the Michigan sample has a reliability index of Cronbach's $\alpha_{MI}=0.55$, compared to Cronbach's $\alpha_{Whole}=0.84$; the reliability of the composite measure of "protecting instructional time" for the Beijing sample is Cronbach's $\alpha_{BJ}=0.58$, compared to Cronbach's $\alpha_{Whole}=0.92$; and the reliability of the measure of "maintaining high visibility" for the Michigan sample is Cronbach's $\alpha=0.59$, compared to Cronbach's $\alpha_{Whole}=0.81$. These differences between the subsamples and

Table 2 Internal consistency of reliability of composite measures of principal learning-centered leadership, Cronbach's α

Variables	Whole Sample	MI	BJ
Framing your school goals	0.77	0.8	0.77
Develop goals that seek improvement over current levels of academic performance			
Frame the school's academic goals in terms of staff responsibilities for meeting them			
Involve school staff in goals development			
Develop goals that are easily translated into classroom objectives by teachers			
Communicating your school goals	0.71	0.69	0.71
Refer to the school's academic goals in informal settings with teachers (e.g., coffee hours)			
Discuss the school's academic goals in formal settings with teachers (e.g., faculty meetings)			
Ensure that the school's goals are reflected in highly visible displays in the school (e.g., poster or bulletin boards)			
Refer to the school's goals in student assemblies			
Supervising and evaluating instruction	0.84	0.55	0.86
Conduct informal observations in classrooms on a regular basis (informal observations are unscheduled, last at least 5 min, and *may* involve written feedback)			
Review student work products when evaluating classroom instruction			
Point out specific strengths in teacher instructional practices in post observation conferences			
Note specific weaknesses of the teacher's instructional practices in written evaluations			
Coordinating curriculum	0.82	0.78	0.83
Make clear who is responsible for coordinating the curriculum across grade levels (e.g., assistant principal, or a teacher)			
Ensure that the school's academic goals are translated into common curricular objectives and classroom activities			
Draw on the results of school wide testing when making curricular decisions			
Assess the overlap between the school's curricular objectives and the achievement test(s) used for program evaluation			
Participate actively in reviewing or selecting curricular materials			
Monitoring students' progress	0.80	0.84	0.66
Distribute test results in a timely fashion			
Inform teachers of the school's performance in written form (e.g., in a memo or newsletter)			
Inform students of the school's performance			
Develop or find the appropriate instructional program(s) for students whose test results indicate a need			
Protecting instructional time	0.92	0.93	0.58
Ensure that instructional time is not interrupted by public-address announcements			
Ensure that students are not called to the office during instructional time			
Ensure that truant students make up lost instructional time			

Table 2 (continued)

Variables	Whole Sample	MI	BJ
Visit classrooms to see that instructional time is used for learning and practicing new skills and concepts			
Cover classes until a late or substitute teacher arrives			
Tutor or provide direct instruction to students			
Maintaining high visibility	0.81	0.59	0.84
Take time to talk with students and teachers during recess and breaks			
Visit classrooms to discuss school issues with teachers and students			
Attend or participate in co- or extra-curricular activities			
Cover classes until a late or substitute teacher arrives			
Providing incentives for teachers	0.79	0.72	0.80
Reinforce superior performance by teachers in staff meetings, newsletters, or memos			
Compliment teachers privately for their efforts or performance			
Acknowledge special effort or performance by teachers in memos for their personnel files			
Providing incentives for student learning	0.77	0.72	0.71
Recognize students who do superior academic work with formal rewards such as an honor roll or mention in the principal's newsletter			
Use assemblies to honor students for their academic work and/or behavior in class			
Recognize superior student achievement or improvement by seeing students in the office with their work products			
Contact parents to communicate improved student performance in school			

The calculation is based on survey data collected by asking principals to rate the extent to which they enacted leadership activities illustrated in this table, by using a four-scale rating: "1=Not at all," "2=Some extent", "3=Moderate extent", and "4=Great extent".

the whole sample in the values of Cronbach's α on these three measures lead us to be cautious with regard to the reliability of these three measures in these three sub-samples. We will acknowledge this limitation in our data analysis and interpretation of results. We excluded other measures of learning-centered leadership behaviors from our final data analysis, such as providing professional development for teachers, because the reliability for those measures was considerably low in one sample (Cronbach's $\alpha < 0.50$).

Moreover, Table 3 reports these composite measures of district evaluation features and associated survey items, as well as their internal consistency. These composite measures include purposes, content foci, sources of evidence, and emphasized leadership activities. The Cronbach's αs are consistently close to or larger than 0.7 for the whole sample and these two subsamples. Furthermore, for each composite measure, there are no significant differences in the magnitudes of the reliability coefficients across these three samples (refer to the whole and the two subsamples). For example, for the first variable of purposes, the Cronbach's $\alpha_{Whole}=0.73$, the Cronbach's $\alpha_{MI}=0.74$ and the Cronbach's $\alpha_{BJ}=0.75$. This consistency indicates a great extent that the items used to aggregate this composite variable of purposes in the two subsamples measured the same latent construct as that in the whole sample.

Table 3 Internal consistency of reliability of composite measures of district principal evaluation, Cronbach's α

	Whole Sample	MI	BJ
Purposes	0.73	0.74	0.75
To provide information for use in making decisions about hiring, principal promotion, (re)assignment, and contract renewal within the district			
To provide evidence for use in determining merit salary increases or sanctions for principals			
To promote the professional development of principals			
To facilitate school restructuring			
To hold principals accountable for student achievement			
Content Foci	0.74	0.80	0.66
Principal leadership behaviors, actions and engagement in activities			
Principals' instructional knowledge and skills			
Principals' school management knowledge and skills			
Principals' organizational impacts on school climate and student achievement			
Sources of Evidence	0.82	0.77	0.69
Observations by supervisor			
Artifacts/actual examples of work provided by principal			
Formal complaints or grievances			
Principal's written evaluations of teachers			
School program audit			
Reports and feedback from teachers, parents, peers, mentors and students			
School wide student achievement data			
Specific Leadership Practices Emphasized	0.85	0.88	0.72
Teacher evaluation activities			
School goal setting			
Meeting Adequate Yearly Progress (AYP)			
Providing professional development programs for teachers			
Curriculum design			
Monitoring student learning during school time			

The raw scales of these items will be included in Tables 5, 6, 7 and 8 in this paper

3.3 Analytic strategies

We first tested the statistical significance of differences between these two subsamples with regard to each key measure of evaluation practice and principal leadership practice. Because many of the district principal evaluation measures were categorical, we used Wilcoxon Rank Sums tests (for rating scale) and Cross-tab Chi-square tests (for dichotomous variables). Since the composite measures of principal leadership practices were continuous, we then performed independent two sample T-tests.

Lastly, we fit two-level Multivariate Hierarchical Linear Models (HMLM) (Raudenbush and Bryk 2002) to the whole sample to examine the extent to which variations in principal leadership activities could be explained by variances in district

principal evaluation practices. Since these measures of district principal evaluation practices were highly correlated, we added one measure at a time to the model.

We also controlled for a series of covariates that may have been confounded with the association between district evaluation and leadership practices. For example, we controlled for principals' working locations by including an indicator of whether the principal was working in Michigan (1=Michigan; 0=Beijing). Within the Beijing sample, we indicated whether the principal was working in Beijing city districts since there were be dramatic differences in school facilities, financial resources, and principal and teacher quality between city schools and rural schools ("1=Within Beijing city" and "0=Outside of Beijing City"). We also controlled for other principal background information, such as years of working experience as a principal, years of working experience as a teacher, gender, educational degree, and educational levels of their schools (Heck and Marcoulides 1996). In sum, the model can be simplified mathematically as:

Level-1 Model

$$
\begin{aligned}
Y_{mi} = {} & (IND1_{mi})\ (\text{Framing school goals}_{1ij})* \\
& +(IND2_{mi})\ (\text{Communicating school goals}_{2ij})* \\
& +(IND3_{mi})\ (\text{Supervising and evaluating instruction}_{3ij})* \\
& +(IND4_{mi})\ (\text{Coordinating curriculum}_{4ij})* \\
& +(IND5_{mi})\ (\text{Monitoring students' progress}_{5ij})* \\
& +(IND6_{mi})\ (\text{Protecting instructional time}_{6ij})* \\
& +(IND7_{mi})\ (\text{Maintaining high visibility}_{7ij})* \\
& +(IND8_{mi})\ (\text{Providing incentives for teachers}_{8ij})* \\
& +(IND9_{mi})\ (\text{Providing incentives for student learning}_{9ij})*
\end{aligned}
$$

$$
Y*_{mi} = \pi_{0i} + \varepsilon_{mi}
$$

Level-2 Model

$$
\begin{aligned}
\pi_{0i} = {} & \beta_{00} + \beta_{01}(\text{Purposes}_j) \\
& \text{Or} + \beta_{01}(\text{Content foci}_j) \\
& \text{Or} + \beta_{01}(\text{Sources of evidence}_j) \\
& \text{Or} + \beta_{01}(\text{Specific leadership practices emphasized}_j) \\
& \text{Or} + \beta_{01}(\text{Frequency}: \text{Being evaluated at least once a year}_j) \\
& \text{Or} + \beta_{01}(\text{Having at least one type of incentive (s) for outstanding leadership} \\
& \qquad\ \text{practices}_j) \\
& \text{Or} + \beta_{01}(\text{Having at least one type of consequence(s) for unacceptable} \\
& \qquad\ \text{leadership practices}_j) \\
& +\beta_{02}(\text{Working in Michigan}_j) \\
& +\beta_{03}(\text{Working in Beijing city}_j) \\
& +\beta_{04}(\text{Years of working experience as a principal}_j) \\
& +\beta_{05}(\text{Working experience as a full} - \text{time teacher}_j) \\
& +\beta_{06}(\text{Being a female}_j) \\
& +\beta_{07}(\text{Held a bachelor and higher degree}_j) \\
& +\beta_{08}(\text{working in an elementary school}_j) \\
& +\beta_{09}(\text{working in a middle or high school}_j) \\
& +r_{0j}
\end{aligned}
$$

Where

Y_{mi} The dependent variables of these nine composite measures of principal leadership activities; m, can be 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, and 9th

IND1   Indicator of the 1st dependent variable
IND2   Indicator of the 2nd dependent variable
IND3   Indicator of the 3rd dependent variable
…
IND9   Indicator of the 9th dependent variable
π_{0i} The intercept at level-1
ε_{mi} The error term at level-1
β_{00} The intercept at level-2
$\beta_{01}-\beta_{09}$ indicates the direction and the strength of the association between each corresponding independent variable and these nine dependent variables simultaneously
r_{0j} The error term at level-2.

4 Results

We first compared learning-centered leadership activities and features of district principal evaluation between Michigan and Beijing principals. We then reported the results from the HMLM analyses to provide evidence on identifying effective evaluation features.

4.1 Comparative results on principal leadership practices

As shown in Table 4, Beijing principals reported practicing several leadership practices to a significantly greater extent than their peers in Michigan, including supervising and evaluating instruction, coordinating curriculum, protecting instructional time, maintaining high visibility, and providing incentives for student learning (see Table 2 for specific items associated with these composite measures). However, there were no significant differences with regard to framing and communicating school goals, monitoring student progress, or providing incentives for teachers.

Table 4 Two sample T-tests of differences in aspects of learning-centered principal leadership practices

	MI	BJ	T-value
Framing school goals	3.590	3.708	−0.86
Communicating school goals	0.339	0.249	1.42
Supervising and evaluating instruction	3.349	3.895	−2.99**
Coordinating curriculum	3.313	3.790	−2.64**
Monitoring student progress	3.336	3.061	1.65
Protecting instructional time	2.922	4.071	−4.98***
Maintaining high visibility	3.184	3.744	−3.03**
Providing incentives for teachers	3.703	3.825	−0.73
Providing incentives for student learning	2.753	3.768	−5.84***

** p-value$\leq$0.05; *** p-value$\leq$0.001

4.2 Comparative results on features of district evaluation

Purposes As included in Table 5, we observed that in both Michigan and Beijing, district principal evaluation was used to provide information for use in making personnel decisions, to promote professional development, and to hold principals accountable for student achievement. But there was less emphasis on using principal evaluation to determine merit salary or sanctions for principals or to facilitate school restructuring. Compared to their counterparts in Michigan, on average, Beijing principals felt that evaluation was more likely to be used to determine merit salary increases or sanctions (p-value ≤ 0.001) or to hold principals accountable for student achievement (p-value $= 0.039 \leq 0.05$).

Content foci There were no significant differences in principals' perceptions of the content emphasized in district evaluations between Michigan principals and their Beijing peers, as demonstrated in Table 6. Across these two locations, the majority of the surveyed principals (ranging from 60% to 81%) indicated that district evaluation focused on their leadership behaviors, their instructional and management knowledge and skills, and their impact on school climate and student achievement.

Sources of evidence Table 7 indicates that compared to Michigan principals, more Beijing principals perceived that leadership evaluation was more likely to be based on

Table 5 Percentage of principals who perceived that the following purposes of principal evaluation were important in their districts by locations

	Not at all Important		A little Important		Somewhat Important		Very Important		Wilcoxon Rank Sums Test
	MI	BJ	MI	BJ	MI	BJ	MI	BJ	
To provide information for use in making decisions about hiring, principal promotion, (re)assignment, and contract renewal within the district	17.11	7.41	21.05	34.57	22.37	41.98	39.47	16.05	0.162
To provide evidence for use in determining merit salary increases or sanctions for principals	71.60	25.00	7.41	32.81	6.17	29.69	14.81	12.50	<.0001
To promote the professional development of principals	13.41	8.54	14.63	29.27	37.80	41.46	34.15	20.73	0.12
To facilitate school restructuring	34.67	11.25	21.33	35.00	29.33	41.25	14.67	12.50	0.056
To hold principals accountable for student achievement	13.92	10.53	16.46	31.58	32.91	40.79	36.71	17.11	0.039

The calculation is based on survey data collected by asking principals to rate how important the purposes of principal evaluation in their districts are, by using a four-scale rating: "1=Not at all important", "2=A little important," "3=Somewhat important extent", "4=Very important"

Table 6 Percentage of principals who indicated that principal evaluation focused on the following aspects by locations

	MI	BJ	Chi-square
Principal leadership behaviors, actions and engagement in activities	80.68	74.71	0.343
Principals' instructional knowledge and skills	67.05	62.07	0.491
Principals' school management knowledge and skills	73.86	77.01	0.629
Principals' organizational impacts on school climate and student achievement	79.55	75.86	0.558

We asked principals to indicate whether district evaluation focused on these aspects listed above in the table. If principal indicated "Yes", we coded the response as "1", otherwise "0"

several sources of evidence, including observation by supervisors; artifacts/actual examples of work provided by principals themselves; formal complaints or grievances from parents and teachers; principals' written evaluations of teachers; school program audits; reports and feedback from teachers, parents, peers, mentors and students; and school wide student achievement data.

Leadership activities emphasized Table 8 shows that a significantly higher percentage of Beijing principals felt that district leadership evaluation emphasized teacher evaluation, provision of professional development programs for teachers, curriculum design, and supervision of student learning during school time than their counterparts in Michigan did. However, there were no significant differences between these two

Table 7 Percentage of principals who perceived that each of the following sources of evidence were emphasized in principal evaluation by locations

	Not at all important		A little important		Somewhat important		Very important		Wilcoxon Rank Sums Test
	MI	BJ	MI	BJ	MI	BJ	MI	BJ	
Observations by supervisor	20	8.11	33.75	14.86	23.75	39.19	22.5	37.84	0.001
Artifacts/actual examples of work provided by principal	23.08	4.35	23.08	18.84	25.64	44.93	28.21	31.88	0.018
Formal complaints or grievances	21.74	9.84	47.83	22.95	15.94	21.31	14.49	45.9	<.0001
Principal's written evaluations of teachers	43.59	15.79	30.77	25	17.95	42.11	7.69	17.11	<.0001
School program audit	51.35	1.25	31.08	8.75	10.81	21.25	6.76	68.75	<.0001
Reports and feedback from teachers, parents, peers, mentors and students	24.36	3.7	28.21	11.11	30.77	28.4	16.67	56.79	<.0001
School wide student achievement data	7.5	0	18.75	3.8	31.25	34.18	42.5	62.03	<.0001

We asked principals to rate the extent to which each of the sources of evidence listed in the table was emphasized in district principal evaluation, by using a four-scale rating: "1=Not at all important", "2=A little important," "3=Somewhat important extent", "4=Very important"

Table 8 Percentage of principals who perceived that various activities were emphasized in principal evaluation by locations

	Not at All Important		A Little Important		Somewhat Important		Very Important		Wilcoxon Rank Sums Test
	MI	BJ	MI	BJ	MI	BJ	MI	BJ	
Teacher evaluation activities	19.75	1.23	32.1	11.11	33.33	40.74	14.81	46.91	<.0001
School goal setting	6.02	3.85	18.07	6.41	27.71	29.49	48.19	60.26	0.053
Meeting Adequate Yearly Progress (AYP)	6.17	1.27	13.58	8.86	25.93	27.85	54.32	62.03	0.177
Providing professional development programs for teachers	12.66	3.8	25.32	16.46	36.71	32.91	25.32	46.84	0.002
Curriculum design	30	5.19	32.5	12.99	23.75	23.38	13.75	58.44	<.0001
Monitoring student learning during school time	12.66	1.39	22.78	11.11	29.11	18.06	35.44	69.44	<.0001

We asked principals to rate the extent to each of the leadership activities were emphasized in their district evaluation by using a four-scale rating: "1=Not at all important", "2=A little important," "3=Somewhat important extent", "4=Very important"

subsamples in terms of evaluating leadership in setting school goals or meeting district-defined achievement requirements.[3]

Other features Table 9 includes comparisons of other features of district principal evaluation. Forty percent of Michigan principals reported that they were evaluated every 3 years and 26.25% were evaluated once a year, while 6% of them reported never being evaluated by district administrators. In contrast, 8% of Beijing principals reported that they had been evaluated every 3 years, and 71.39% of them had been evaluated once a year, while none of them reported never being evaluated.

Furthermore, compared to their Michigan peers, a significantly larger percentage of Beijing principals indicated that there were consequences for them if they were evaluated as outstanding or unacceptable ($p \leq 0.05$). Forty percent of the principals in Beijing indicated that there was at least one type of consequence for the principal if she/he was evaluated as outstanding (including confidential recognition or award, bonus, advancement on salary scale, promotion to another building, promotion to district central office, or public recognition or award). However, only 10% of principals in Michigan indicated that their district offered at least one of these awards if they were evaluated as outstanding. Thirty nine percent of the Beijing principals indicated that there was at least one type of consequence if the principal was evaluated as unacceptable (such as dismissal, demotion to assistant principal or other position, reassignment to another building, pay cut, denial of pay increase, or removal from key district decision making processes), while only 23% of principals in Michigan indicated facing at least one of these consequences.

---

[3] There was no exact term of "Adequate Yearly Progress (AYP)" in Beijing. But Beijing Department of Education had similar practices of monitoring student progress at the school level as AYP in Michigan.

Table 9 Percentage of principals who perceived other features of principal evaluation by locations

In general, how often have you been evaluated during your career as a principal?	MI	BJ	Chi-square Test
Every 3 years	40	8.7	
Every other year	15	2.9	
Once a year	26.25	71.01	
Twice a year	12.5	17.39	
Three or more times each year	0	0	
Never	6.25	0	
the consequences for the principal if she/he is evaluated as outstanding? (confidential recognition or award, bonus, advancement on salary scale, promotion to other building, promotion to district central office, public recognition or award)	10.23	40	<.0001
the consequences for the principal if she/he is evaluated as unacceptable? (such as dismissal, demotion to assistant principal or other position, reassignment to other building, pay cut, denial of pay increase, removal from key district decision making processes)	23.86	38.89	0.031

4.3 Association of features of district evaluation with principal learning-centered leadership practices

After we descriptively examined the various leadership practices and features of district evaluation, we combined Michigan and Beijing samples to explore the extent to which variations in principal learning-centered leadership activities were associated with variation in district principal evaluation practices. Results are included in Table 10. We found that principals' perceptions of several features of district leadership evaluation were associated with their enactment of a number of learning-centered leadership activities (all *p-values* of evaluation features ≤0.001). These features include evaluation purposes, content foci, sources of evidence, and leadership activities

Table 10 HMLM estimation results of the association of district principal evaluation with principal leadership practices

	Coefficient Estimates
Purposes	0.208*** (0.037)
Content Foci	0.479*** (0.032)
Sources of Evidence	0.283*** (0.045)
Specific Leadership Practices Emphasized	0.271*** (0.038)
Frequency: Being evaluated at least once a year	0.049 (0.049)
Having at least one type of incentive(s) for outstanding leadership practices	0.050 (0.042)
Having at least one type of consequence(s) for unacceptable leadership practices	0.019 (0.040)

Covariates include: Working in Michigan, working in Beijing city, years of working experience as a principal, years of working experience as a full-time teacher, being a female, held a bachelor and higher degree, working in an elementary school, and working in a middle or high school

** *p*-value≤0.05; *** *p*-value≤0.001

emphasized in district principal leadership evaluation. Specifically, principals were more likely to engage in learning-centered leadership when they felt that districts were using evaluation a) to make decisions regarding hiring, promotion, re(assignment), and contract renewal; b) to provide evidence for use in determining merit salary increases or sanctions; c) to promote their professional development; d) to facilitate school restructuring; or e) to hold them accountable for student achievement. Moreover, when district principal evaluation focused on leaders' instructional and management knowledge and skills, leadership behaviors, and organizational impacts, principals were more likely to engage in various learning-centered leadership activities.

Furthermore, greater diversity in the sources of evidence used in evaluation was positively associated with more frequent engagement in learning–centered leadership. These sources included those collected and provided by multiple actors, including district administrators, principal themselves, teachers, parents, peers, and even students. Finally, when evaluation emphasized curriculum, instruction, and student learning, principals were more likely to exercise influence in these areas. However, neither the frequency of evaluation (e.g., being evaluated at least once a year) nor the positive or negative consequences of evaluation had a significant impact on levels of implementation of learning-centered leadership activities.

5 Discussion

5.1 Substantive interpretations of results

Our data in this study revealed that Beijing principals placed more emphasis on curriculum, instruction, and direct interactions with students than their counterparts in Michigan. Such differences may stem from the fact that a higher percentage of Beijing principals had more than 3 years of teaching experience than Michigan principals. In China, almost all K-12 public school principals are former senior teachers who have demonstrated pedagogical expertise in classrooms, and principals in China are respected as head teachers. Moreover, Chinese principals continue to have teaching responsibilities and consistently improve their teaching knowledge and skills to pursue higher-level teaching certification (Yang 2006). Principals often teach example lessons and allow other teachers to critique and model their instruction (Chu and Liu 2010). Therefore, they have expertise in designing classroom activities, selecting teaching materials, and interacting with students to motivate them to perform at a higher level.

In Michigan, however, only slightly more than half (62%) of the surveyed principals had more than 3 years of full-time teaching experience. Principals without multiple years of teaching experience may still be able to manage their schools with regard to framing and communicating school goals, distributing test results, and providing incentives for teachers, but they may not be proficient at directly participating in teaching, conducting formal and informal observations in classrooms on a regular basis, providing feedback on teachers' instructional practices, covering classes for teachers, ensuring that instructional time is used for learning new skills and concepts, or directly interacting with students about learning particular subjects. Using Stein and Nelson (2003)'s framework to interpret these similarities and differences between

Michigan and Beijing principals, we can tentatively conclude that they acted similarly with regard to the outer layer of instructional leadership in supporting teaching and learning (such as setting goals, distributing test results, and providing incentives for teachers), while Beijing principals were more likely to engage in the core activities of instructional leadership (such as direct interactions with students about their learning, observing classrooms, and providing feedbacks to teachers) than their Michigan peers.

With regard to our descriptive analysis of district evaluation features, principal evaluation in both Beijing and Michigan shared the same purposes with regard to personnel management and development. However, we found that Chinese principals perceived that districts were more likely to use evaluation a) to hold them accountable for student achievement and b) for purposes of merit pay increases or sanctions. These findings are not surprising, given the long tradition in China of accountability centered on student learning outcomes, the fact that Chinese principals have long been evaluated and compensated as governmental officials, and the fact that official rankings of principals are determined by the status of their schools, which in turn determines principals' income and their future professional careers (Chu and Yang 2009; Qian 2009). The position of principals in the hierarchical structure of school governance and the high-stakes nature of principal evaluation may also lead to the finding that although both Michigan and Beijing principal evaluations generally addressed the same leadership content, Chinese principals perceived a stronger impact of specific aspects of district evaluation (e.g., in terms of both sources of evidence and specific leadership behaviors emphasized in evaluation) than their Michigan peers.

Finally, this study found that after controlling for multiple covariates, including their school location, years of principalship and full-time teaching experience, gender, and educational levels of their schools, when a principal perceived a higher level of emphasis of district evaluation on purposes, content, sources of evidence, and specific principal leadership activities, he/she was more likely to focus on learning-centered leadership activities. These aspects of evaluation provide relatively detailed information about districts' expectations regarding principal leadership and also indicate the ways in which principals can meet these expectations in schools. In contrast, frequency of evaluation did not explain the variance in leadership behaviors, nor did receiving at least one type of award or punishment based on evaluation. A possible explanation for this is that timing and incentives enact pressure on principals but contain little specific guidance on what actions principals should take. However, future studies are needed to provide further evidence regarding how principals can best be motivated and supported.

5.2 Limitations

This study has several limitations. First, we only collected data from a sample of principals in Michigan and a sample of principals in the Beijing metropolitan area. The sample from Beijing, to a large extent, represents general practices of principal leadership and district evaluation due to the centralized and uniform school governance system in China. As a matter of fact, Chen's study (2011) indeed showed no differences in principals' instructional leadership activities across geographical locations in

China. However, the Michigan sample may not represent principal leadership activities and district evaluation across the U.S. Therefore, the generalizability of the findings in this study is limited.

Second, we did not implement a large-scale pilot study of our survey instrument before we administered it in Michigan or Beijing. Therefore, the internal validity of our findings may be violated by the possibility that we used a weak instrument. However, the measures in the instrument were developed based on a thorough literature review of leadership practices that lead to gains in student learning and a comprehensive literature review of features of evaluation practices. The fact that the Cronbach's α values on all composite measures were sufficiently high serves as another indicator of the reliability of our measures.

Third, our data analyses were performed on cross-sectional data. Therefore, we do not intend to draw causal inferences regarding the effect of district evaluation on principal leadership practices; rather, we are attempting to draw attention to effective features of district principal evaluation and their possible associations with principal leadership behaviors. Moreover, we are not advocating for the effectiveness of the Chinese leadership and principal evaluation system because we acknowledge profound differences in cultures, social values, and political governance structures between China and the U.S.; rather, we intend to call attention to how we can learn from other countries' practices of school governance and leadership that may address educational challenges in the U.S. with regard to improving instructional quality and learning outcomes.

5.3 Policy implications

Despite the limitations, this study provides valuable empirical evidence to support the policy assumption that district principal evaluation can be effectively used to communicate school leadership responsibilities and serve as a powerful policy instrument to influence principals' behavior (Catano and Stronge 2006). To serve these purposes, district evaluation on principal leadership should emphasize the ways in which principals can impact teaching and learning outcomes. Although the surveyed Michigan districts and principals utilized these activities as their Beijing counterparts did, they should place significantly greater emphasis on principals' leadership in developing teachers, directly participating in instruction, and directly interacting with students about their learning activities and outcomes.

District principal evaluation alone may not lead to desired learning-centered leadership activities unless it is supported by other principal personnel management strategies. For example, when recruiting principals, having at least 3 years of full-time teaching experience can be listed as one of the key qualifications. We also recommend connecting principal evaluation with professional development for school leaders. Evaluation can serve as a diagnostic tool to identify the needs for principal professional development. We further suggest that besides management skills and knowledge, professional preparation and training programs for principals should emphasize effective instructional strategies and equip them with sufficient and updated instructional knowledge and skills to allow them to have meaningful interactions with teachers and students about complex subject matter.

6 Conclusion

Principal evaluation has become a key component of national policy debates regarding how to develop effective leaders. To contribute to these debates, this study draws on survey data to explore how principals in Michigan and metropolitan Beijing behaved differently in enacting learning-centered leadership and how district principal evaluation can contribute to the variation in principal leadership practices. Empirical evidence from these two regions indicates that district evaluation should emphasize principals' responsibilities for promoting instructional and learning outcomes in determining the purposes, contents, and sources of evidence used in principal evaluation. Designing systematic evaluation and support systems for effective school leadership includes recruiting high-quality principal candidates, collecting data on principals' instructional leadership responsibilities, and using data to inform principal professional development.

References

Bottoms, G. (2001). *What school principals need to know about curriculum and instruction*. New York, NY: The Wallace Foundation.

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

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: The University of Chicago Press.

Burch, P., & Spillane, J. P. (2003). Elementary school leadership strategies and subject matter: reforming mathematics and literacy instruction. *The Elementary School Journal, 103*(5), 519–535.

Catano, N., & Stronge, J. H. (2006). What are principals expected to do? Congruence between prinicial evaluation and performance standards. *NASSP Bulletin, 90*(3), 221–237.

Chen, F. (2011). Empirical investigation on instructional leadership behaviors of elementary and secondary school principals. *Shanghai Jiaoyu Keyan, 3*, 34–36.

Chu, H. (2003). Construct the knowledge base for principal professionalization. *Theory and Practice of Education, 23*(12), 27–32.

Chu, H., & Cravens, X. C. (2011). *Principal professional development in China: Challenges, opportunities, and strategies*. Nashville, TN: Vanderbilt University.

Chu, H., & Liu, J. (2010). Promoting principal instructional leadership. *The Elementary and Middle Schools Administration, 3*, 4–16.

Chu, H., & Yang, H. (2002). Principal's specialization and its guarantee system. *Theory and Practice of Education, 22*(11), 20–26.

Chu, H., & Yang, H. (2009). *Towards principal professionalization*. Shanghai: Shanghai Education Publishing House.

Coburn, C. E. (2001). Collective sensemaking about reading: how teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170.

Cravens, X. C. (2008). *The cross-cultural fit of the learning-centered leadership framework and assessment for Chinese principals*. (Unpublished doctoral dissertation). Vanderbilt University, Nashville, Tennessee.

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*, 197–228.

Davis, S. H., & Hensley, P. A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education, 13*(4), 383–403.

Doud, J. L., & Keller, E. P. (1998). *The K–8 principal in 1998: A ten-year study*. Alexandria, VA: National Association of Elementary School Principals.

Goldring, E., Porter, A. C., Murphy, J., Elliott, S. N., & Cravens, X. (2007). *Assessing learning-centered leadership: Connections to research, professional standards, and current practices*. New York, NY: The Wallace Foundation.

Goldring, E., Cravens, X. C., Murphy, J., Elliott, S. N., Carson, B., & Porter, A. C. (March, 2008). *The evaluation of principals: What and how do states and districts assess leadership?* Paper presented at the annual meeting of American Education Research Association, New York, NY. Retrieved from http://peabody.vanderbilt.edu/Documents/pdf/LSI/AERA_EvaluationPrincipals.pdf.

Guo, K. (2005). The trend of evaluation for elementary and secondary school principals. *Instruction and Administration*, (1), 5-7.

Hallinger, P. (2011). Leadership for learning: lessons from 40 years of empirical research. *Journal of Educational Administration, 40*(2), 125–142.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: a review of empirical research, 1980–1995. *Educational Administration Quarterly, 32*(1), 5–44.

Hallinger, P., & Heck, R. (1999). Next generation methods for the study of leadership and school improvement. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration, second edition* (pp. 141-162). San Francisco: Jossey-Bass.

Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal, 96*(5), 527–549.

Heck, R. (1992). Principals' instructional leadership and school performance: implications for policy development. *Educational Evaluation and Policy Analysis, 14*(1), 21–34.

Heck, R. H., & Marcoulides, G. A. (1996). The assessment of principal performance: a multilevel evaluation approach. *Journal of Personnel Evaluation in Education, 10*(1), 11–28.

Heck, R. H., Larsen, T. J., & Marcoulides, G. A. (1990). Instructional leadership and school achievement: validation of a causal model. *Educational Administration Quarterly, 26*(2), 94–125.

Hipp, K. A. (1996). *Teacher efficacy: Influence of principal leadership behavior.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Jing, W. (2010). *School leadership in two countries: Shared leadership in American and Chinese high schools.* (Unpublished doctoral dissertation). Arizona State University, Arizona.

Kimball, S. M., Milanowski, A., & McKinney, S. A. (2007). *Implementation of standards-based principal evaluation in one school district: First year results from randomized trial.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Lashway, L. (2003). *Improving Principal Evaluation. ERIC Digest. Access ERIC: Full Text* (071 Information Analyses–ERIC IAPs No. EDO-EA-03-09). Oregon: ERIC Clearinghouse on Educational Management, Eugene, OR.

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly, 30*(4), 498–518.

Leithwood, K. (1996). School restructuring, transformational leadership and the amelioration of teacher burnout. *Anxiety, Stress and Coping, 9*, 199–215.

Leithwood, K., & Janzi, D. (2005). A review of transformational school leadership research. *Leadership and Policy in Schools, 4*, 177–199.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*: The Wallace Foundation.

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher, 29*(2), 4–16.

Louis, K. S., Marks, H. M., & Kruse, S. D. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33*, 757–798.

Mangin, M. M. (2007). Facilitating elementary principals' support for instructional teacher leadership. *Educational Administration Quarterly, 43*(3), 319–357.

Marcoulides, G. A., Larsen, T. J., & Heck, R. H. (1995). Examining the generalizability of a leadership model: issues for assessing administrator performance. *International Journal of Educational Management, 9*(6), 4–9.

Marks, H. M., & Nance, J. P. (2007). Contexts of accountability under systemic reform: implications for principal influence on instruction and supervision. *Educational Administration Quarterly, 43* (1), 3–37.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.

Murphy, J. (2007). *Restructuring through learning-focused leadership*, pp. 71–84. Handbook on Restructuring & Substantial School Improvement, edited by Herbert J Walberg. Center on Innovation and Improvement, U.S. Department of Education.

Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007a). Leadership for learning: a research-based model and taxonomy of behaviors. *School Leadership and Management, 27*(2), 179–201.

Murphy, J. F., Goldring, E. B., Cravens, X. C., Eliot, S. N., & Porter, A. C. (2007b). *The Vanderbilt Assessment of Leadership in Education: Measuring Learning-Centered Leadership*. Nashville, TN: Vanderbilt University.

Newmann, F. M. (1997). How secondary schools contribute to academic success. In K. Borman & B. Schneider (Eds.), *Youth experiences and development: Social influences and educational challenges*. McCutchan: Berkeley, CA.

Newmann, F. M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education, 108*(4), 259–299.

Newmann, F., Smith B., Allensworth, E., & Bryk, A. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis,23*(4), 297-321.

OECD. (2010). *PISA 2009 at a Glance*: OECD Publishing. http://dx.doi.org/10.1787/9789264095298-en.

Payne, D., & Wolfson, T. (2000). Teacher professional development–the principal's critical role. *NASSP Bulletin, 84*(618), 13–21.

Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Qian, H. (2009). *Secondary school principalship in China: Leading at the cusp of change.* (Unpublished doctoral dissertation). The Chinese University of Hong Kong, Hong Kong, China.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Reeves, D. B. (2005). *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. Thousand Oaks, CA: Corwin.

Stein, M. K., & D'Amico, L. (2002). Inquiry at the crossroads of policy and learning: A study of a district-wide literacy initiative. *Teachers College Record, 104*(7), 1313–1344.

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

Stein, M. K., & Spillane, J. P. (2005). How can educational leaders support and promote teaching and learning? New conceptions of learning and leading in schools. In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research in educational leadership* (pp. 28–45). New York: Teachers College Press.

Stine, D. O. (2001). Developing an evaluation system to improve principal performance and accountability. Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Sun, M. & Youngs, P. (2009). How does district principal evaluation affect learning-centered principal leadership? Evidence from Michigan school districts. *Leadership and Policy in Schools, 8*(4), 411–445.

Sun, M., Frank, K., Penuel, W. R., & Kim, C. (2010). *Formal versus informal leaders: How institutions penetrate schools*. Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Sun, M., Penuel, W. R., Frank, K. A., Youngs, P., & Gallagher, H. A. (2011). *Shaping professional development to promote the diffusion of instructional expertise among teachers*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Walker, J., & Slear, S. (2011). The impact of principal leadership behaviors on the efficacy of new and experienced middle school teachers. *NASSP Bulletin, 95*(1), 46–64.

Xin, Z., & Wang, L. (2006). Review of elementary and secondary school principals evaluation. *Theory and Practice of Education, 26*(9), 13–15.

Xu, P. (2010). Improving principals' instructional leadership through reviewing and evaluating classroom instruction. *The Elementary and Middle Schools Administration, 03*, 14–16.

Yang, H. (2006). Principal career ladder: promoting elementary and secondary school principals' professionalism. *Modern Elementary and Secondary Education, 146*(4), 63–66.

Zhao, Q., & Liu, J. (2010). Principals' instructional leadership in China. *The Elementary and Middle Schools Administration, 03*, 10–13.