The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0957-8234.htm



Received 14 September 2012
Revised 21 January 2013
Accepted 27 January 2013

# The cross-cultural validity of the Learning-Centered Leadership framework and its assessment instrument for principals

Xiu Chen Cravens
*Peabody College of Education and Human Department, Vanderbilt University, Nashville, Tennessee, USA*

## Abstract

**Purpose** – The purpose of this paper is to select a theoretical framework for effective school leadership that is connected with research, standards and current practices in the USA, and examine its validity and generalizability cross-culturally.

**Design/methodology/approach** – The paper uses both qualitative and quantitative methods through expert panel evaluation, cognitive interviews, and field testing of the instrument. The author asks: How well does the Learning-Centered Leadership (LCL) framework align with the professional standards and current practices of principals in urban Chinese schools in the opinion of the experts? Is there evidence that its leadership assessment instrument has construct validity in Chinese urban schools based on the re-examination of its content and measurement criteria? And is there evidence that the instrument is yielding consistent results when taken by the intended participants? How effective are the analytic strategies employed by this study in detecting the equivalences and discrepancies in how educational leadership is defined and evaluated, between two vastly different educational systems?

**Findings** – The paper finds evidences that give support to the claim that there is strong cross-cultural alignment on the overarching goal of improving student learning. However, the US framework and assessment will need to be modified to reflect the Chinese reform priorities that emphasize the balance between academic and social learning.

**Originality/value** – The belief that there are common elements in contemporary international educational policy has brought growing interest in sharing leadership theories and successful models cross-culturally. This paper addresses the challenges in understanding the complexity of cross-cultural translation of theories and applications, and explores viable solutions to meaningful adaptation.

**Keywords** Assessment, Cross-cultural comparison, Evaluation, School leadership, Validation study

**Paper type** Research paper

## Introduction

The belief that there are general and common elements in contemporary, international education policy (Ball, 1998; Brown and Lauder, 1996) has brought growing interest in sharing leadership theories and successful models of effective schools cross-culturally. Such interests are particularly strong from countries that are at the beginning stage of establishing a knowledge base for school leadership development (Chu, 2003; Hallinger, 2011; Hallinger *et al.*, 2005; Walker *et al.*, 2012).

The author thanks Hongqi Chu, Ellen Goldring, Andrew Porter, Joseph Murphy, Steven Elliott, James Guthrie, Hong Wang, and Nianbo Dong for their thoughtful comments on earlier drafts of this paper. The author also gratefully acknowledges the funding support provided by the Vanderbilt International Office and data collection assistance provided by South China Normal University.

Journal of Educational Administration
Vol. 52 No. 1, 2014
pp. 2-36
© Emerald Group Publishing Limited
0957-8234
DOI 10.1108/JEA-09-2012-0102

المنارة للاستشارات
www.mana

In China, the role of educational leadership, especially the role of school principals is getting significant attention from policy makers and educational administration scholars. A movement of principal professionalization that aims at developing newer and stronger leadership is gaining momentum. However, two main challenges exist for principal leadership development in China: to provide training for preparation and growth, and to implement systemic accountability to support such efforts (Chu, 2003; Zhao and Wang, 2007). For historical and political reasons, domestic theories and empirical evidence for effective educational leadership are limited in China. A variety of imported school leadership theories and models have been introduced to the Chinese research community (Walker *et al.*, 2012), including strands such as transformational leadership (Leithwood and Jantzi, 1993), instructional leadership (Hallinger and Murphy, 1987), distributed leadership (Spillane, 2006), and moral leadership (Sergiovanni, 1992), to name a few. The challenge, however, is to understand the complexity of cross-cultural translation of educational theories, their applications (Astiz *et al.*, 2002) and furthermore, to get beyond the recognition of such challenge by mapping out possible pathways to sustainable and meaningful cross-cultural adaptation (Dimmock and Walker, 2000; Hallinger and Leithwood, 1998; Heck, 1996).

In this paper, we showcase an empirical study that explores one of such cross-cultural pathways to establish a knowledge base for school leadership development. We select a theoretical framework for effective school leadership that is connected with research, standards and current practices in the USA, and explores its validity and generalizability in urban Chinese schools. The purpose is threefold: first, to empirically test the theoretical fit of the Learning-Centered Leadership (LCL) framework (Porter *et al.*, 2006) when compared with the professional standards and current practices in urban Chinese schools; second, to examine the empirical evidence for validity and reliability of the principal assessment instrument developed from the LCL framework; and third, to extrapolate replicable methodological strategies that may be effective in addressing challenges in the cross-cultural transfer of leadership theories and their operational instruments.

## Background and literature

### *The changing role of school principals in China today*

Contemporary Chinese educational reform largely reflects the nation's economic development and societal change. From the early 1980s to the turn of the century impressive progress was made, especially in the implementation of nine-year compulsory education, the elimination of illiteracy, and the enhancement of higher education and vocational training (National Center for Education Development Research (NCEDR), 2008). In the 1990s, quality-oriented education (*su zhi jiao yu*) started as an initiative promoting holistic student development both academically and socially, and gathered more momentum into the early twenty-first century (Gu, 2010; State Council of the Chinese Central Government, 1993, 1999). Quality-oriented education introduced a blueprint for promoting education that highlighted the cultivation of independent thinking, creative spirit, and the practical ability of the nation's young generation. Improving education quality, and at the same time balancing education efficiency with equity became the new policy imperatives (Chu, 2008; NCEDR, 2008). More importantly, the new priority brought on important changes to teacher preparation programs, curriculum structure and content, instructional methods, and most relevant to this paper, the ways of measuring school success and school leader effectiveness.

المنارة للاستشارات

www.mana

The belief that loyalty and conformity with the hierarchical order of the society are fundamental virtues is deeply imbedded in the Chinese value system, which has profound influence on how leaders in China think, operate, and receive their appraisals. School principals, for example, have long been classified as government officials and must align their decisions and actions with political authorities (Li, 2004). They are also held accountable more as moral leaders than any other aspect of the leadership responsibilities (Gao *et al.*, 2006).

Such long-held beliefs, values and ideologies, however, are being challenged more than ever. At the center of educational reform in China, the role of school principals has been a focal point of discussion and research interest in recent years (Chu and Cravens, 2012; Walker *et al.*, 2012). Identifying the core functions of school principals and developing training strategies are considered the corner stone of the reform efforts in response to the rapidly changing social and economic demands (Chu, 2003; Li, 2004).

The system that treated school principals as typical government officials has been widely criticized by Chinese educational policy researchers in more recent years, particularly after the establishment of a teacher certification system in 1993 as the contrast between the teaching profession and the role of school principals became more observable. The research community points out that the existing system has several detrimental effects on school management. First, principals have the strong tendency to treat the schools as an extension of the government and manage the school with top-down approaches. Second, the authoritarian nature of the principal position creates an atmosphere in the school community that gives little consideration to student-centered instructional focus. Third, because only seniority matters most in the bureaucratic hierarchy, principals rely on their prior experiences as teachers to manage their schools, without additional training and support to enhance their knowledge, ability and techniques as school leaders (Huang, 2004).

Traditionally, China's education was very highly centralized. Since 1985, the Ministry of Education gradually delegated more authority to local governments and school districts. Schools were encouraged to involve their communities and principals became more directly involved in management decisions. However, more authority also means more obligations, more responsibilities, and more accountability. The decentralized school system requires principals to have strong leadership in making a broad range of decisions.

As the Chinese society gradually opened to the global market and with the momentum of economic reform, the tight control over the education system was being relaxed. The new priorities of quality-oriented education, the decentralization of governance, the need for diversified financial resources, and the ever intensifying competition among schools and their students for access to higher education and the job market, combined, are pushing the redefinition of the role of school principals to the top of the educational reform agenda (Chu, 2003; Chu and Cravens, 2012).

*A new knowledge base for school leadership*

As principal leadership development gaining more attention in the larger context of national educational reform, the need to identify a framework to improve both the quality and equity of public education became a pressing priority. Moreover, advocates for principal leadership development called for systemic support for the effort of professionalize school principals in China (Chu, 2003), which include: a comprehensive human resource management structure for principals; a principal licensure system;

المنارة للاستشارات

www.mana

training for principal preparation and professional development; assessment and evaluation; and incentive and promotional opportunities for principals.

As an integral part of the systemic support, leadership assessment is a significant "condition of leadership" impacting leaders' behavior (Chu and Cravens, 2012). However, there has been little evidence of effective practices of using formative or summative assessments to measure and develop leadership knowledge and skills (Zhao and Wang, 2007). In fact, there has not been any specific governmental stipulation regarding principal evaluation except for a recommendation made by the Ministry of Education (MOE, 1992), titled *The Draft Opinions on Enhancing the Development of School Principals in the Nation*, in which four dimensions of assessing principal effectiveness were proposed: values, abilities, diligence, and achievement (MOE, 1992). These four dimensions have been widely used at provincial, city, and township levels as the guidelines for principal evaluations. However, how the four dimensions are operationalized and which formats are used vary greatly (Zhao and Wang, 2007).

Despite the significant growth of research in the area of educational administration, studies focussing on principal evaluation have been few (Chu and Cravens, 2012; Zhao and Wang, 2007). Among the limited number of published research articles on principalship in Mainland China, a majority of them limit their topics to introducing theories and practices of other countries or stop at recounting the needs and issues in school leadership development without branching into specifics. In their review of the literature on Chinese principalship written in English and Chinese between 1998 and 2008 (Walker *et al.*, 2012), they found that a majority of the pieces were non-empirical and focussed on prescriptions and commentaries. They also found that while there were an increasing number of empirical studies addressed imported frameworks, indigenous investigations, and contextual influences, they lacked rigor and generalizablity for reasons that might be "traced to an underdeveloped research infrastructure, an inadequate knowledge base, a centralized ideology, and the enduring influence of traditional values" (p. 18).

Overall, research and practices in principal leadership assessment suffer from two deficiencies: the lack of a sound theoretical framework for principal leadership that links the objectives of education with leadership standards; the lack of an assessment system that is developed with empirical research evidence and is valid and reliable (Chu, 2003; Zhao and Wang, 2007).

*LCL framework and the Vanderbilt Assessment of Leadership in Education™ (VAL-ED)*

The connection between learning-focussed leadership and leadership assessment as it contributes to educational outcomes has been more intensely examined in recent decades in the USA under policies that emphasize systemic accountability and improvement (Goldring *et al.*, 2009a; Portin *et al.*, 2006). Research indicates that when designed appropriately, executed in a proactive manner, and properly implemented, it has the power to enhance leadership quality and improve organizational performance at three levels: at the individual level, assessment can be used as a benchmarking tool for essential personnel functions such as documentation for annual reviews and compensation. At the level of continuous learning and development, leadership assessment can serve as a powerful communication tool, providing both formative and summative feedback to a school leader, where incumbent school principals may make informed decisions regarding development and improvement by identifying gaps between existing practices and desired outcomes. At the level of collective

المنارة للاستشارات

www.mana

accountability for school-wide improvement, leadership assessment can set the organizational goals and objectives for the school leader (Goldring *et al.*, 2009a). When the domains of school leadership that impact student achievement are included as the assessed targets (Goldring *et al.*, 2009b; Heck *et al.*, 1990; Heck and Marcoulides, 1996), leadership assessments help school leadership focus on those behaviors that are associated with student learning.

Any leadership evaluation model that tries to capture all of the subtleties of the principal's role, and operationalize all of the day-to-day activities of the principal is doomed to fail. A more realistic question is: how can we measure the most important indicators of effective school leadership related to school performance?

Building upon the conviction that instructionally focussed and change-oriented leadership are especially effective frames for education, the LCL framework was established to inform the crafting of a new evaluation system for school leaders and school leadership teams (Murphy *et al.*, 2007). A comprehensive review of largely US-based research literature (see Goldring *et al.*, 2009b; Murphy *et al.*, 2007) reveals two key dimensions of highly effective leadership related to student learning and achievement: core components and key processes (see Appendix 1 for definitions). Core components refer to what principals or leadership teams must accomplish to improve academic and social learning for all students, while key processes refer to how leaders create and energize those core components. Effective LCL, according to this framework, is at the intersection of the two dimensions – core components created through key processes (Table I).

The VAL-ED requires respondents to make judgments about a principal's leadership behaviors that influence teachers' performance and students' learning. Respondents are asked how effective the principal is at *s*pecific actions that affect the core components and key processes of LCL. The effectiveness ratings range from 1 = ineffective to 5 = outstandingly effective for each of 72 behaviors (see Appendices

| | Key processes | | | | | |
|---|---|---|---|---|---|---|
| Core components | Planning | Implementing | Supporting | Advocating | Communicating | Monitoring |
| High standards for student performance | | | | | | |
| Rigorous curriculum (content) | | | | | | |
| Quality instruction (pedagogy) | | | | | | |
| Culture of learning and professional behavior | | | | | | |
| Connections to external communities | | | | | | |
| Systemic performance accountability | | | | | | |

**Source:** Porter *et al.* (2006)

**Table I.**
Learning-Centered Leadership framework conception chart

www.mana

2.1 and 2.2). These behaviors sample all 36 cells of our conceptual model of leadership equally and thus serve as indicators of the construct of leadership the LCL framework desires to measure (Elliott, 2008).

Based on the LCL framework, the VAL-ED was designed to be markedly different from current instruments employed by states and districts throughout the USA in that, the VAL-ED uses 360 degree feedback, from teachers, principals, and supervisors; the content of the assessment focusses on LCL that leads to increases in student achievement; the assessment is of leadership behaviors, not knowledge, dispositions, or personal characteristics of leaders; the VAL-ED requires respondents to identify evidence on which they are basing their assessment of principal behaviors; and the psychometric properties are clearly documented.

The VAL-ED has gone through a series of qualitative and quantitative instrument pilot studies and field testing (Porter *et al.*, 2010a, b). As of the spring of 2011, the VAL-ED is being used by more than 280 school districts with about 2000 schools nationwide for principal evaluation purposes. Studies that collect further psychometric evidence for validity and reliability are ongoing supported by the US Department of Education. In January 2008, *Education Week* reported that researchers with learning point associates, a non-profit educational consulting firm based in suburban Chicago reviewed eight principal-performance instruments being used by school districts and concluded that VAL-ED comes closest to measuring the leadership attributes and behaviors that research finds to be associated with how well students perform (Maxwell, 2009). VAL-ED also was rated the best among the instruments for validity and reliability, meaning that the assessment measures what it is supposed to measure and yields consistent results by the same report.

The well-documented validation process of the VAL-ED, both theoretically and empirically, makes it an attractive candidate for studying the cross-cultural transfer process, where viable tools for educational leadership assessment are few and far between internationally (Cravens, 2012; Murphy *et al.*, 2011). In fact, western theories and school leadership models have been gradually introduced to China and were frequently referenced since the late 1990's (Chen, 2004; Du, 2004). While the imported leadership theories have been widely discussed and referenced in the Chinese research literature (Walker *et al.*, 2012), players active in education reform are still in search for a framework that incorporates effective theories that can be operationalized into practical tools for leadership development in the local context.

However, there are significant differences in national education policy context and reform imperative between the two systems. In the USA the core challenge facing America's schools, especially urban schools, is improving overall student performance on core subjects and decreasing the achievement gap (Linn, 2000; National Governors Association, 2008; Peterson *et al.*, 2011). School leadership, especially principal instructional and transformational leadership, is regarded as an essential driving force to implement processes and conditions such as rigorous academic standards, high-quality instruction, and a culture of collective responsibility for students' academic success (Goldring *et al.*, 2009b). While the LCL framework conceptually defines the focus of leadership effort as student learning, both academically and socially (Murphy *et al.*, 2007; Porter *et al.*, 2006), its theoretical and empirical rationale (Goldring *et al.*, 2009b) is largely grounded in research that links effective leadership to student outcome such as in-school achievement (e.g. grades on common final exams), performance at exit from school (e.g. graduation), and more distal indices of accomplishment (e.g. college graduation). The weight, as it appears in the US context, is relatively heavy on academic learning.

المنارة للاستشارات

www.mana

Interestingly, a reverse trend appears to be emerging in China. Decades of emphasis on standardized testing to promote academic achievement is giving grounds to both governmental mandates and societal demand for graduates that are well-rounded citizens, equipped with not only book knowledge but moral values, creative mindset, mental and physical health, and versatile interests (Cravens *et al.*, 2011; Zhou, 2004).

Could the LCL framework and the VAL-ED contribute to building a new knowledge base in China? We have three specific research foci: first, the fit of the theoretical framework – How well does the LCL framework, conceptualized by core components and key processes, align with the professional standards and current practices of principals in Chinese schools in the opinion of the experts? Second, the validity and reliability of the VAL-ED scores – is there evidence that the instrument has construct validity, based on the examination of its content and measurement criteria? And is there evidence that the instrument is yielding consistent results when taken by the intended participants? Third, the viability of identifying replicable procedural elements to test the cross-cultural fit of leadership theories and instruments – How effective are the analytic strategies employed by this study in detecting the equivalences and discrepancies in how educational leadership is defined and evaluated, between two vastly different educational systems?

## Method and data

### *Methodological challenges of validating conceptual constructs across cultures*

While a majority of previous research on cross-cultural validation has been in the clinical psychology and medical field, numerous attempts have also been made to examine the construct equivalence in management and leadership concepts (Dorfman and Howell, 1997; Hallinger and Kantamara, 2001; Heck, 1996; Heck and Marcoulides, 1996). Studies show that not only may the particular leadership framework being emphasized vary culturally, but the same framework may have different meaning within different cultures (Heck and Marcoulides, 1996). Psychometric theorists point out those sources of error or invalidity can be organized into three broad categories (Hambleton *et al.*, 2005): cultural/language differences, technical issues in the design, and interpretation of results. We address each area of concern by enlisting specific strategies that will be included in the study design.

*Cultural and language differences*. Four elements in the assessment process – construct equivalence, test administration, item formats used, and the influence of speed on examinee performance – are highlighted as the main areas where cultural and language differences may affect test results (Hambleton *et al.*, 2005).

First, determining whether construct equivalence exists between two cultures involves judgmental strategies, which may involve interviewing or observing people from the cultures of interest, researching the cultures of interest, and asking others who know about the cultures. Second, communication problems during text administration can pose a serious threat to the validity of test results. For example, there might be special problems with understanding the rating scales of the VAL-ED because the two-dimensional nature of item cells. This concern could be addressed by ensuring the translation and presentation of the framework, its conceptual elements, and how the instrument's items reflect the conception are clear through an iterative improvement process. The proper selection of test administrators could be helpful. Principals and teachers might not be willing to fill out the survey when administrators cannot explain fully the purpose of the survey and give clear instructions. Third, there are also concerns over the differential familiarity with particular item formats that may

المنارة للاستشارات
www.mana

present another source of invalidity of test results in cross-cultural studies. For example, although selected response items such as multiple-choice items have been used extensively in the USA for achievement testing, it cannot be assumed that everyone is as familiar with multiple-choice items. Lastly, while the VAL-ED is not timed, how much time teachers, principals and their supervisors have to fill out the survey and the possible bias caused by rushing, say, at the end of an in-service meeting, should be considered during the study.

*Technical issues in the design*. Technical factors that may influence the validity of tests adapted for use in other languages and cultures may include the test itself, selection and training of translators, the process of translation, and the judgmental designs for adapting tests (Hambleton *et al.*, 2005). The VAL-ED translation is to be done by bilingual researchers who know the cultures well, especially the target culture (i.e. the culture associated with the language of the adapted test) to ensure that the nuances and subtleties of a subject area can be lost on a translator unfamiliar with the subject matter. The translation of the VAL-ED is to go through the back-translation process known as the best and most popular of the judgment designs (van Widenfelt *et al.*, 2005), where two translators adapt the VAL-Ed from the source language (English) to the target language (Chinese). Two different translators take the Chinese VAL-ED and adapt it back to English. Then, the original and the back-translated versions of the VAL-ED are compared and judgments are then made about their equivalence by the translating team.

*Interpretation of results*. Cross-cultural studies should not be used to support arguments about the superiority or exceptionality of nations as if the international comparative study is the equivalent of a horse race with winners and losers (Westbury, 1992). At best, these studies provide only a "snapshot" of differences that exist, and provide only a limited basis for interpreting the results. In this context, to gain a better understanding when interpreting scores, other relevant factors external to the tests or assessment measures and specific to a nationality are to be considered. Leadership assessment, for example, is embedded in educational policies and standards, wealth, standard of living, cultural values, and so on, which may all be essential factors for properly interpreting the results across the USA and Chinese settings.

### *Analytic strategies and data*

This study aims to ensure that construct equivalence is established between the original VAL-ED instrument and the translated version by appropriately choosing judgmental designs (such as back-translation), validity and reliability measures, and statistical analyses to provide data bearing on the question of item and test equivalence across language and cultural groups.

We took a three-step approach to answer the research questions and address the methodological concerns specifically related to the cross-cultural adaptation of assessment instruments (Hambleton, 2002). We capture this approach in Table II, which maps our analytical strategies and sources of validity evidence.

Specifically, in Step 1, we employed an expert panel to examine the face validity and content validity of the LCL framework and the VAL-ED. Participants of the expert panel ($n = 12$) were comprised of five faculty members of education administration from two major teaching universities, two officials of provincial bureau of education, and two principals in principal professional development training. This purposive sample was selected to identify individuals who were knowledgeable in the subject area and could provide valuable feedback regarding current professional standards

www.mana

| Psychometric properties across settings | Analytic strategies | Sources of validity evidence | Sampling and data collection |
|---|---|---|---|
| Fit of the theoretical framework | Alignment analysis (qualitative) | In-depth survey results and narrative feedback from expert panel on the importance and relevance of the assessment items (face and content validity) | Step 1<br>Purposive sampling based on area expertise |
| Validity and reliability of the instrument | Cognitive lab interviews (qualitative) | "Think aloud" responses from intended users on their interpretation of the assessment items (content validity) | Step 2<br>Purposive sampling based on role match with intended assessment target |
| | Factor analysis (quantitative) | The empirical cluster pattern of the assessment items as compared with the conceptual structure (construct validity) | Step 3<br>Representative sampling or convenience sampling with representative key characteristics, of intended assessment target |
| | Internal consistency (quantitative) | Consistency of the assessment item scores across subscales, users, and overall (reliability) | |
| | Performance nomination (quantitative) | Relationship with other measures used concurrently in the intended setting (criterion validity) | |

**Table II.**
Analytic strategies and sources of validity evidence

المنارة للاستشارات

www.mana

and practices of school principals in China. The panel respondents were asked to perform a content alignment analysis and also examine the quality of instrument translation. Each member was given an alignment rating form (Appendix 3) and asked to examine the LCL framework and VAL-ED assessment items using a Likert response scale from two angles: reality – the level each principal leadership behavior described in the VAL-ED was practiced in Chinese urban schools day-to-day, and importance – the level of attention the same leadership behavior should receive. To ensure that the existing LCL framework and VAL-ED could be tested for cross-cultural fit, all 72 items for the core components and key processes were translated to maintain the original construct meaning. Based on results from the alignment analysis, translation of the instrument items was to be refined for better clarity and improved reflection of the construct.

In Step 2, we conducted cognitive interviews to further probe into the instrument content construction. The cognitive lab interviews were designed to as a qualitative measure to ensure that respondents were interpreting the VAL-ED items as they were originally intended and the full range of appropriate responses could be captured. Typically, the cognitive interview requires respondents to "think aloud" as they work through a questionnaire, providing the researcher with a play-by-play of their cognitive processes. This process was also intended to yield important insight regarding the extent to which construct equivalence was achieved after the LCL framework elements (core components and key processes) and with items translated into Chinese. A purposive sample of respondents was identified from participants of principal and teacher training programs offered at a major teaching university in the fall of 2007. A total of six participants (two principals, two teachers, and two supervisors of principals) from urban schools participated in the interviews.

In Step 3, we examined the validity and the reliability of the instrument quantitatively using rating scale scores collected from a sample of urban schools. The refined version of the VAL-ED in Chinese was administered to principals, their principals, and teachers in a set of schools located in a provincial capital in southern China. Analyses of the VAL-ED results included descriptive statistics, correlations, factor analysis, estimates of internal reliability, and comparing the VAL-ED results with the Chinese evaluation criteria.

Identifying schools to participate in the study was carefully planned. To avoid possible coercion from the local bureau of education (comparable to the US district superintendency), efforts were made to recruit and consent principals prior to recruiting and consenting their supervisors. The principals were identified from the 50 participants of principal training programs in a southern metropolitan city in the fall of 2007. The purpose of the research project was fully explained in the invitation letter, along with issues of risks, benefits, confidentiality, and the voluntary nature of participation. The voluntary principals were also asked to let the researcher know if his/her supervisor could be contacted to participate in the study. From the 35 principals that volunteered for participation, 20 schools were selected based on their representativeness of academic ranking, economic condition of the school zone, school size, and school type (elementary, middle, high) in mind. The supervisors were contacted separately via letter and follow-up telephone calls. Participating school teachers were sent a recruitment letter separately explaining the purpose of the research project and the studies, along with issues of risks, benefits, confidentiality, and the voluntary nature of participation. The final sample included 1,165 teachers from 19 schools, 18 principals from the same 19 schools, and six supervisors for the 18 principals.

المنارة للاستشارات

www.mana

Table III provides a summary of the profiles of the schools that participated in the study. Although the 19 schools were not selected through a randomized process, the sample included a good range of schools in the city in terms of school type, percentage of senior teachers, geographic location and school size. There were seven elementary school (37 percent), three high schools (16 percent), three middle schools (16 percent), four schools that were have both middle and high school grades (21 percent), and two vocational schools (grades 10-12). Teacher classification was based on years of experience and instructional quality – percentage of senior teachers at each school is an important indicator of not only the seniority of the faculty, but also student achievement in Chinese schools (Chen, 2006). Among the 19 schools, percentages of senior teachers ranged from as low as 10 percent to as high as 90 percent. The schools also varied in size, from small elementary schools with fewer than 500 students to large high schools of more than 2,000 students. One vocational school had more than 5,000 students. The schools were located in six different school districts among the ten that were in the city. Some of the districts were in the downtown area, and some were in the outskirt of the city considered very rural and agricultural less than 15 years ago. Many of the student families were migrant workers who worked in the city in these newly transformed schools. The principals from the sampled schools also varied in their experience and gender. Ten principals were male and nine were female, with an average experience as a principal of 5.3 years, ranging from less than one year to more than ten years. The principals had an average of 3.7 years of tenure at the existing schools.

The overall return rate for the teachers was 97.4 percent, and school-level teacher return rates were above 90 percent for all the schools in the sample. The return rate for the principals was 95 percent, with one principal out of 19 did not return the assessment. The return rate for the supervisors was 100 percent.

## Results

*Face validity and content validity – detesting the "reality gap"*

Feedback from the expert panel shed light on how content elements of the LCL framework and its assessment instrument – the VAL-ED – were aligned with the practice and standards in urban Chinese schools. The reality dimension gauged the extent to which the leadership behaviors of the VAL-ED weighed in the workload of Chinese school principals – the higher the weight (rated on a scale of 1-5), the better the alignment of the VAL-ED with current principal practices. The importance dimension demonstrated the extent to which the leadership behaviors were believed to be important to the success of the schools and students – the higher the rating (also on a scale of 1-5), the better the alignment of the VAL-ED with the participants' definition of effective leadership. The difference between the ratings given to importance and reality, is the perceived "reality gap". This alignment approach provided three sets of comparison measures on the fit of the VAL-ED framework and items: how the VAL-ED was aligned with the current practices; how the VAL-ED was aligned with the respondent's definition of LCL, and how the practice might be different from such definition.

Table IV shows that the reality ratings for the six core components and the six key processes are all below 3.5. Among them, connection with external communities and systemic performance accountability received ratings lower than 3.0, a level that was below average; quality instructions received a 3.46 average, which was the highest. On the other hand, the core components all received ratings close or above 4.0 on

المنارة للاستشارات
www.mana

| 学校 ID | School type | No. of teacher assessments issued | No. of teacher assessments returned | Teacher assessment return rate (%) | Total no. of Students | Percentage of senior teachers | Years as principal | Years as principal of this school |
|---|---|---|---|---|---|---|---|---|
| 1 | Elementary | 65 | 59 | 90.77 | 1468 | 71 | 2 | 2 |
| 2 | Elementary | 36 | 35 | 97.22 | 669 | 48 | 5 | 5 |
| 3 | Elementary | 41 | 40 | 97.56 | 1012 | 66 | 2 | 2 |
| 4 | Elementary | 44 | 43 | 97.73 | 1193 | 90 | 2 | 2 |
| 5 | Elementary | 68 | 67 | 98.53 | 780 | 53 | 2 | 2 |
| 6 | Elementary | 55 | 55 | 100.00 | 1629 | 60 | 12 | 6 |
| 7 | Elementary | 30 | 30 | 100.00 | 436 | 45 | 10 | 2 |
| 8 | High | 60 | 56 | 93.33 | 2783 | 46 | 7 | 7 |
| 9 | High | 153 | 149 | 97.39 | 2686 | 32 | 5 | 2 |
| 10 | High | 33 | 33 | 100.00 | 2200 | 32 | 7 | 6 |
| 11 | Middle | 51 | 45 | 88.24 | 1174 | 10 | 9 | 9 |
| 12 | Middle | 52 | 51 | 98.08 | 2742 | 27 | na | na |
| 13 | Middle | 47 | 47 | 100.00 | 1000 | 10 | 3 | 2 |
| 14 | Middle and High | 36 | 35 | 97.22 | 780 | 13 | 6 | 5 |
| 15 | Middle and High | 121 | 120 | 99.17 | 1355 | 38 | 3 | 2 |
| 16 | Middle and High | 35 | 35 | 100.00 | 783 | 13 | 6 | 2 |
| 17 | Middle and High | 155 | 155 | 100.00 | 4500 | 21 | 6 | 4 |
| 18 | Vocational | 87 | 83 | 95.40 | 5300 | 20 | 8 | 8 |
| 19 | Vocational | 27 | 27 | 100.00 | 2000 | 15 | 1 | 1 |

**Table III.**
Profiles of sampled schools

www.mana

| *Core component* | *High standards* | *Rigorous curriculum* | *Quality instruction* | *Culture of learning* | *External communities* | *Performance account* |
|---|---|---|---|---|---|---|
| Reality | 3.27 | 3.07 | 3.46 | 3.34 | 2.42 | 2.98 |
| Std | 0.56 | 0.63 | 0.76 | 0.73 | 0.72 | 0.71 |
| Importance | 4.36 | 3.97 | 4.40 | 4.22 | 3.74 | 4.10 |
| Std | 0.30 | 0.30 | 0.41 | 0.41 | 0.58 | 0.44 |
| Difference[a] | 1.09 | 0.9 | 0.94 | 0.88 | 1.32 | 1.12 |
| *Key process* | *Planning* | *Implement* | *Supporting* | *Advocating* | *Comm* | *Monitoring* |
| Reality | 3.26 | 3.23 | 3.45 | 2.66 | 3.02 | 2.92 |
| Std | 0.53 | 0.68 | 0.79 | 0.56 | 0.65 | 0.66 |
| Importance | 4.22 | 4.14 | 4.28 | 3.97 | 4.19 | 4.00 |
| Std | 0.38 | 0.41 | 0.39 | 0.31 | 0.51 | 0.24 |
| Difference[a] | 0.96 | 0.91 | 0.83 | 1.31 | 1.17 | 1.08 |

**Notes:** [a]Paired *t*-test for the difference between the group means is significant at $p < 0.001$, $df = 5$

**Table IV.**
Average ratings for reality and importance of the VAL-ED items

average on importance, indicating that the leadership behaviors were important to the success of schools and students in the eye of the expert panel but might not have been carried out to the same extent in practice. When the 72 items were sorted by the key processes, the same could be said about the difference between the mean scores for reality and for importance. All six processes received mean scores for relevance at below 3.50. The highest mean score is for supporting at 3.45, and the lowest is for advocating, at 2.66. The mean scores for the importance of the six processes, however, were mostly above 4.0, except for advocating at 3.97. The paired *t*-test on reality vs importance show a correlation of 0.92 between the two sets of mean scores for the core components, with the mean of importance ratings (4.13) higher than the mean of reality ratings (3.08) at a highly statistically significant level ($p < 0.001$, $df = 5$). For the key processes, the correlation is 0.90 between the mean ratings for importance and reality and the two sets of ratings are different at a statistically significant level ($p < 0.001$, $df = 5$).

The fact that the mean ratings for reality were consistently lower than the importance ratings for all six core components and all six key processes of the VAL-ED indicated the presence of gaps between what were considered necessary to enhance learning-centered school results and what were believed to be practice by principals.

The variation of the "reality gap" among the components and processes was an indication that the fit of the VAL-ED might be better in some areas than the others. Better alignment appeared to be in the core components related to curriculum and instruction, but not related to working with external communities or in the process of advocating for student needs. Such views were supported by both the practitioners and the researchers on the expert panel.

Analysis of the reality and importance ratings was also done at the item level. The ten items with the largest "reality gap" between current practice and perceived importance indicate that the inadequacies appeared to concentrate on two areas: working with families and community, and attending to the needs of students that have difficulties in learning or from diverse backgrounds.

The "reality gap" was further noted by the panel members' written comments. The two principals commented on Items 19, 20, and 43 (check Appendix 2.2 for all items), which address the needs of students either lagging behind academically or from diverse backgrounds: "The standardized curriculum makes it difficult for the principal

المنارة للاستشارات
www.mana

and teachers to attend to the individual needs of students." "It is only possible to help students who lag behind to reach basic proficiency but not to the higher standards in the current environment." "Diversity among Chinese students is not prevalent therefore not considered a priority." "Addressing diversity would require large amount of resources and energy that we current cannot afford." Regarding connecting with family and the community, practitioners and researchers on the panel commented that "outreach to external entities it is not common practice for schools." "Information flow from school to parents tends to be one-way. Parents' input is not well taken into consideration."

On the other hand, some leadership behaviors were perceived as being practiced at levels desirable by their importance. Among the ten items as having the least differences between the reality and importance ratings, four (40 percent) items were related to the core component of rigoros curriculum. For example, Item no. 37 was the only leadership behavior that had a negative "gap" between reality and importance (average importance = 4.08, average reality = 4.50), meaning what was practiced was beyond perceived importance. Such negative gap was reported by both the practitioners (−0.25) and the researchers (−0.50).

Comments from the panel indicated that there might be too much discipline and order were in practice and more than what were perceived as necessary. "There is little room for school-level or classroom-level decision making for curriculum content but to comply with the mandates, thus the relevance and importance ratings are high," commented a researcher. One superintendent noted: "There is too much emphasis on test scores in urban Chinese schools, especially on written examinations." The other superintendent pointed out that item 16 did not apply for all schools in China (implements a rigorous curriculum in all classes), because the new government policy called for an end to the tracking practices during the nine-year compulsory education period, but the practices were still allowed in high schools.

Overall, the panel members did not believe that significant expansion of the framework was necessary and the current conception largely reflected the leadership domain that addresses student learning. While the overall ratings indicated that the leadership behaviors of the VAL-ED core components and key processes aligned well with the desired leadership behaviors for student learning, the same elements appeared to have weaker alignment with the current practices of school principals.

The existence of the "reality gaps" and the variation in the gaps draw attention to the multiple-faceted nature and the complexity of determining content validity, as new policies are being formed nationally and implemented locally. The sizes of the "reality gap" varied among the six core components, and the perceptions on the gap were slightly different between the practitioner and the researcher groups. Moreover, the panel members gave nuanced explanations to why the sizes of "reality gaps" existed and varied in the national education reform context. We see such results directly linked with a policy life cycle to "support the continuous development and improvement of educational policy and practice from the individual level and site level to the system level" (Cheng, 2002, p. 14). Reviews of major large-scale education evaluations in the USA found large between-school variation in implementation and have shown that the within-school variation in implementation is often as great as or greater than variation between schools (e.g. Desimone, 2002; Lee and Bryk, 1989). Studies on reform implementation in the Asia-Pacific region also identified significant discrepancies between national policy intentions and local implementation strategies (Hirosato, 2001; Ng, 2006).

المنارة للاستشارات

www.mana

*Instrument content construction – reaching a nuanced understanding*

The validation of the instrument content featured cognitive interviews with two officials of a local bureau of education, two principals, and two teachers, in a provincial capital in urban school settings. The two officials were directors of their school districts and had overall supervisory responsibilities over the school principals in the districts. One school principal was from an elementary school, and the other was from a middle school. The two teachers were from an elementary school.

The interview protocols were modeled after the process used in the US validation work for the VAL-ED (Porter *et al.*, 2010a). Each participant was asked to review the 72 items as if he/she was assessing a real principal. The only problem encountered in applying the protocols was that the Chinese participants found it very hard to think aloud during the interview. This was particularly true with the school directors. Speaking one's mind without reservation, especially as someone of authority, is out of the cultural norm in China. In many cases, to the point of item by item, the interviewer had to probe how the item was perceived, whether it was easily understood, and if there were any places that were ambiguous or did not apply.

During the review of VAL-ED items, periodically participants were stopped and asked when giving a rating, if he/she was thinking of the quality or the impact of the behavior or the frequency of the behavior. In most cases, the participants answered quality or impact, except for items related to connections with external communities where several participants said because related activities were so rare, frequency was a factor in their rating decisions.

To specifically address how the VAL-ED items were aligned with the new educational reform initiatives, the interviewer asked each respondent to read the definition for the core component of high standards of student learning and asked probing questions regarding their understanding of the definition: What is your understanding of high standards? What are the standards that you are thinking of? What is high? What is low? What is your understanding of student learning? What does learning entail in your mind? Consistently the participants emphasized that it was important to include the concept of "developing" and not just the focus on studying for high test scores but, they added, that the latter was more of a reality for Chinese schools and students. They pointed out that while the coexistence of the two aspects might be the intent, the ratings might appear to be high only because academic achievement goals were clearly established at the district level with designated administrative staff who monitored school progresses. However, this might not be a true reflection on how schools were doing on social learning, a new educational reform mandate by the Ministry of Education for quality-oriented education.

On items for the core component of high-quality instruction, participants consistently commented that principals were usually too busy to have intimate knowledge of what happened during classroom teaching. In Chinese urban schools, the tasks of working with teachers on instruction such as evaluating how instructional time was used (VAL-ED Item 35) typically fall within the responsibility of one of the assistant principals or the instruction director. Participants believed that ratings were likely to be high because coaching, team teaching, and lesson planning were well developed and highly organized practices in Chinese school. In other words, the effectiveness of the principal at ensuring the school achieves high-quality instruction could be reached by working with administrators assigned to lead instructions and teachers' collaborations. They also mentioned that in most case, principals did not have the authority to make hiring decisions for new or transferring teachers with the exception of few newly established

المنارة للاستشارات
www.mana

school districts had piloted school-based hiring practices, therefore items regarding teacher hiring might not be applicable.

The respondents felt that the items regarding culture of learning and professional behaviors were a bit too generic. The Chinese schools had a long tradition of encouraging teachers to work together. The work team could be formed either by subject or by the common group of students that the teachers teach. Typically teachers of different subjects were assigned to a grade of students and would follow the same grade throughout the middle school or high school period. Even in elementary schools, it was common to follow the grade for two to three years. The practices of mentoring for novice teachers, lesson planning, and peer-learning were considered the norm, according to the participants. However, some participants pointed out that these elements, even though considered important to the "culture of learning" specifically for the teaching faculty, were covered minimally among the VAL-ED items, while most of the items focussed on a culture of learning for the students. On the other hand, one school director talked about there was indeed a disconnect between school culture building and student learning in a full sense that also encompassing all-around development. "Too much focus is put on raising test scores," he said, "there might be calls from the government and from the district level for moral education as a priority for school culture building, but the principals find it hard to follow."

Compared with other core components, connections with external communities was an area that the participants perceived as receiving the least attention from schools and their principals. "Most of our schools operate within the school doors," several participants said. Practices such as building business partnerships to support social and academic learning (Item 52) were rare except for vocational schools, for example. "I am forecasting an average of 3 for the principals that you will assess next for this category," a school director predicted. However, some forms of communications with the external community did exist, and there was an increasing push from parents and the government reform policies for more connections with families and society in general. For example, the school directors gave examples of new initiatives at the district level such as the promotion of the "Three New Excellence" program that aimed at helping students to become excellent citizens in the society, excellent students at school, and excellent children at home. According to the participants, communications with families and the community often were generated organically from the bottom up by teachers or parents, but support and monitoring mechanisms were usually missing.

Participants reported that the items for the last core component, systemic performance accountability, could be easily understood. Systemic accountability as a term has a direct translation counterpart in Chinese. One of the teachers commented that this term was still relatively new to the education system and might not be familiar to all the teachers.

Among the six key processes of the VAL-ED, planning, implementing, and monitoring appeared to be less problematic to the participants. In comparison, supporting, advocating, and communicating received more questions and comments.

For the key process of supporting, participants pointed out that some of the VAL-ED items might have assumed that principals had certain level of authority and flexibility in providing financial resources for learning-related activities, which was not the case in most Chinese urban schools. Budgets and expenditures for Chinese urban schools were based on set formulas and the principals had little control over the funds allocated for specific purposes. Providing support for building the learning

المنارة للاستشارات

www.mana

culture financially (item 41), for example, might be difficult, according to the interviewed participants.

Items related to the key process of advocating describe leadership behaviors that act on behalf of the diverse needs of students. These items were identified by the participants of the interviews as "more of a wish than practical reality" in Chinese urban schools. For example, Items 7 and 8 measure the effectiveness of the principal at ensuring that the school attends to students with special needs. On these two items both school directors commented that though it might be the intention of the principals and teachers, it would be difficult to implement with large class sizes (at least 50 students per class) and the pressure of standardized testing. They also suggested changing "special needs" to "students having difficulties in learning" to be more applicable in the Chinese setting. For Item 19 (advocating a rigorous curriculum that honors the diversity of students and their families), school directors and principals commented that while it might be encouraged by the principals for teachers to creatively use the curriculum to better fit students' individual needs, it would not be realistic to expect this leadership behavior at the school level using the word "ensuring" due to the standardized curriculum content and the achievement testing structure. Similar comments were made by the two interviewed teachers on Item 31 (advocating for all students to regularly experience effective instruction) that students who lagged behind academically might not get the same quality of teaching as the advanced students that were in different academic tracks. This was particularly true for students in their last year in middle or high schools facing placement examinations.

Respondents reported that communicating as a key process was generally clear. However, the respondents also noted that items measuring both the key process of communicating and the core component of connections with external community were the least practiced in their schools.

*Reliability and validity of the VAL-ED scores – evidence from the field*

*Missing data*. The validity and reliability examination of VAL-ED score results was carried out for 19 schools using the final version of the VAL-ED assessment in Chinese. The final sample included 18 principals, 1,165 teachers, and six school directors that were the supervisors of the 18 principals. Data were examined in terms of missing pattern and missing frequency. High percentages of missing data could be threats to construct validity and be caused by various reasons: poor translation, sensitive or uncomfortable subject, inapplicable scenarios, or confusion over the question. We found that missing data occurred but the problem appears to be relatively small in size and with largely normal distributions cross the core components and key processes. For the teacher assessment, 15 percent of the observations missed more than 10 percent of the effective rating scores. The missing data rates for principal and supervisor assessments were lower than 10 percent per observation across the board. The missing scores were also checked at the item level. The missing-data percentages for the 72 items hovered around the average of 9 percent, with a standard deviation of 2 percent, ranging from the lowest of missing 6 percent (Item 42) to the highest of missing 15 percent (Item 60). The distribution of the item-level missing-data percentages were quite normal. Another important criterion for data quality for the VAL-ED scores is the percent of "don't know" as a response for the effectiveness rating. This percent is very low for all the VAL-ED scales at 1-2 percent of the total observations. The only exception is the core component of connections to external communities with a 4.5 percent missing rate, possibly a reflection of the phenomenon

المنارة للاستشارات
www.mana

reported by the expert and cognitive interview participants that working with external communities was less practiced by the principals.

*Factor structure*. Exploratory factory analysis (EFA) was conducted to examine the factorial validity of the VAL-ED items in the Chinese setting. In other words, we aimed to check if the core components and key processes posited in the VAL-ED theoretical framework could be supported by the observable data. Two empirical rotation methods were compared to see which is more suitable for this study. The orthogonal rotation method is the most common rotation of any kind and focusses on maximizing the differences among the pattern/structure coefficients of factors (Thompson, 2004). Another less common rotation method, the oblique rotation, can be used to evaluate the data structure when the factors are expected to be correlated (Thompson, 2004). Both methods were performed. The results were very similar in terms of number of factors identified and the common variances accounted for. Comparatively, the orthogonal method yielded much clearer and interpretable factors than the oblique method. Therefore a principal-axis factor analysis, rotated using Kaiser's (1958) varimax criterion was used to examine the 72-item instrument.

A cutoff of 0.40 was used to interpret solution from the orthogonal rotation (Hair *et al.*, 1998). Items that loaded high on one factor and relatively low on all the others were marked in Table V, whereas items with low loadings and/or dual-factor loadings were not listed. Among the 72 VAL-ED items, 59 (82 percent) of them had factor loadings larger than 0.40 and 13 do not. The 59 items loaded onto seven factors, three of which had eigen values larger than 1.0, accounting for 87.7 percent of the cumulative (total) variance. The other four factors accounted for another 6.0 percent of the total variance. Together, the seven factors accounted for 93.7 percent of the total variance.

The EFA results showed a factor structure that reflected the theoretical framework of six core components only to a limited extent. The most noteworthy discrepancy was the lack of discerning power of the items to reflect the six-component framework. The common variances of the items were concentrated on three main factors. Among the 59 items that had significant factor loadings, 26 of them were concentrated on one factor explaining 81.9 percent of the total variance (also see Table IV). There were only two other factors that had eigen values larger than 1.0. The eigen value is the sum of the squared correlations of each variable and the factor identified. Eigen value larger than 1.0 is generally considered the cutoff for a meaningful factor (Stevens, 1996). As illustrated in the scree plot (Figure 1), when the eigen value gets close to 1.0, the plot level levels off to a linear decreased pattern.

To better understand how the factor structure based on the empirical data fit the VAL-ED framework, factor loadings larger than 0.40 were retained and plotted into the two-dimension LCL framework matrix for core components and key processes, numbered by their factor clustering concentration level from 1 to 7 (Figure 2). We see that the first four core components were more distinctively represented by the factor structure. The clustering of items around the principal factors was by no means clean-cut and some "bleeding" existed among the four components. The items clustering around Factor 1, the largest factor, were concentrated in the last two core components, connections to external communities and systemic performance accountability, including all 12 items for Component 5 and seven items for Component 6. Although the factor structure did not show strong evidence of clustering of items by key processes, the last two key processes, communicating and monitoring, both share the common variances that were also marked by Factor 1. This very well may be an indication of the "bleeding" of the these key processes with the core components of connections with

المنارة للاستشارات

www.mana

| Factor 1 | | Factor 2 | | Factor 3 | | Factor 4 | | Factor 5 | | Factor 6 | | Factor 7 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item no. | Eigen | Item no. | Eigen | Item no. | Eigen | Item no. | Eigen | Item no. | Eigen | Item no. | Eigen | Item no. | Eigen |
| 67 | 0.42 | 44 | 0.43 | 17 | 0.43 | 1 | 0.50 | 70 | 0.50 | 34 | 0.48 | 39 | 0.48 |
| 68 | 0.42 | 33 | 0.45 | 11 | 0.44 | 2 | 0.50 | 71 | 0.58 | 36 | 0.53 | 40 | 0.49 |
| 10 | 0.43 | 31 | 0.45 | 16 | 0.44 | 3 | 0.67 | 72 | 0.59 | 35 | 0.55 | 38 | 0.54 |
| 65 | 0.43 | 26 | 0.47 | 12 | 0.47 | 4 | 0.66 | | | | | | |
| 69 | 0.43 | 27 | 0.50 | 15 | 0.48 | 5 | 0.46 | | | | | | |
| 45 | 0.44 | 42 | 0.51 | 7 | 0.48 | | | | | | | | |
| 66 | 0.44 | 28 | 0.54 | 25 | 0.51 | | | | | | | | |
| 64 | 0.46 | 30 | 0.66 | 19 | 0.52 | | | | | | | | |
| 63 | 0.46 | 29 | 0.69 | 14 | 0.53 | | | | | | | | |
| 48 | 0.47 | | | 13 | 0.54 | | | | | | | | |
| 43 | 0.48 | | | | | | | | | | | | |
| 47 | 0.49 | | | | | | | | | | | | |
| 41 | 0.50 | | | | | | | | | | | | |
| 32 | 0.51 | | | | | | | | | | | | |
| 50 | 0.54 | | | | | | | | | | | | |
| 49 | 0.56 | | | | | | | | | | | | |
| 57 | 0.57 | | | | | | | | | | | | |
| 58 | 0.58 | | | | | | | | | | | | |
| 52 | 0.62 | | | | | | | | | | | | |
| 59 | 0.66 | | | | | | | | | | | | |
| 55 | 0.67 | | | | | | | | | | | | |
| 51 | 0.69 | | | | | | | | | | | | |
| 56 | 0.69 | | | | | | | | | | | | |
| 60 | 0.70 | | | | | | | | | | | | |
| 53 | 0.72 | | | | | | | | | | | | |
| 54 | 0.73 | | | | | | | | | | | | |
| Total eigen | | | | | | | | | | | | | |
| 42.56 | | 1.80 | | 1.22 | | 0.93 | | 0.86 | | 0.69 | | 0.63 | |
| Total variance | | | | | | | | | | | | | |
| 81.92% | | 85.40% | | 87.74% | | 89.53% | | 91.18% | | 92.51% | | 93.71% | |

**Table V.**
Factors identified by exploratory factor analysis

**Notes:** Extraction method – principal axis factoring – was used. For clarity, only values equal to or greater than 0.40 are provided. Among the 72, 13 items have loadings lower than 0.40 and are not listed in the table. They are Items 6, 8, 9, 18, 20, 21, 22, 23, 24, 37, 46, 61, 62

external communities and systemic performance accountability, echoing the feedback given by the expert panel members and the cognitive interview respondents.

The exploratory factor analysis lets the data tell the story and provides a map of the construct domain based on the effectiveness ratings of the teacher assessment from 19 schools and 1,165 teachers. Encouragingly, without being confined to a priori factor structure, the empirical results found preliminary evidence for the first four VAL-ED core components to stand as unique factors that were sufficiently distinctive from one and another to embody valid constructs individually and combined for LCL. Meanwhile, noteworthy discrepancy between the theoretical framework of the VAL-ED and findings did exit. Fewer factors underlying the data were identified than the framework hypothesizes, evidenced by the clustering of only three main factors accounting for nearly 90 percent of covariance. Furthermore, the factor structure was less than clear. The item clustering patterns appeared to be scattered across components and processes, and too many items clustered around the core components

المنارة للاستشارات

www.mana

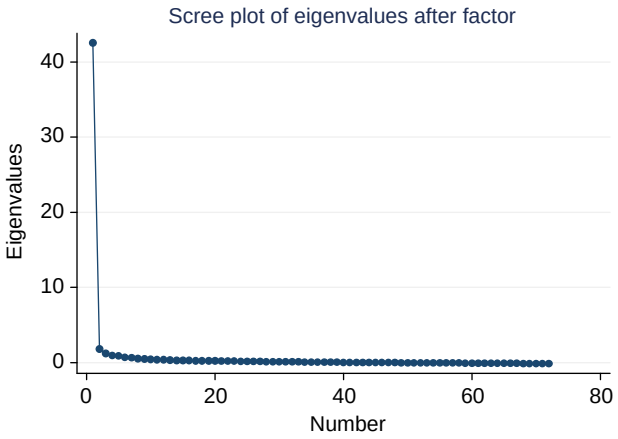



**Figure 1.**
Scree plot of eigen values for factor determination

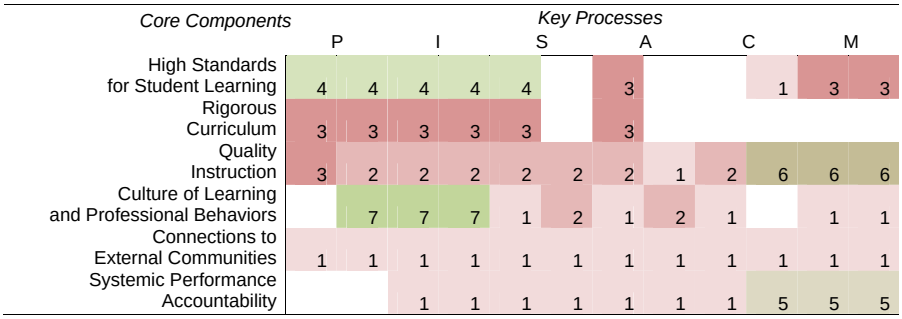



**Figure 2.**
Plotting factor structure with Chinese data in the existing LCL framework

**Note:** Factor loadings larger than 0.40 were retained and plotted into the two-dimension LCL Framework matrix for core components and key processes, numbered by their factor clustering concentration level from 1 to 7 (listed in Table IV)

of connections with external communities and systemic performance accountability. The findings indicate that modification and refinement to the items and even possibly to the original framework would be necessary to reach a better fit between the actual factor structure and the theoretical conception.

*Internal consistency.* Reliability is an essential part of any test or assessment and concerns the consistency of the scores. Two methods were used to estimate the reliability of VAL-ED scores. First, Cronbach's $\alpha$'s (Cronbach, 1951; Cronbach and Meehl, 1955) of each scale for core components and for key processes were calculated, and then inter-rater reliability – teacher-teacher, teacher-principal, teacher-supervisor, principal-supervisor – were determined.

The internal consistency of the VAL-ED items was first examined with the teacher VAL-ED scores for the Cronbach $\alpha$ values on the subscales of core components and key processes. The results are presented in Table VI. Based on the teacher assessment effectiveness ratings, all subscales exhibited excellent internal consistency with Cronbach's $\alpha$'s larger than 0.90 and most of them at about 0.95. The $\alpha$ value for the 72-item overall scale was 0.96.

المنارة للاستشارات

www.mana

| Scale | Cronbach's α | Mean score | Std | SEM | CI-low | CI-high |
|---|---|---|---|---|---|---|
| High standards | 0.946 | 4.29 | 0.64 | 0.15 | 4.05 | 4.53 |
| Rigorous curriculum | 0.947 | 4.21 | 0.65 | 0.15 | 3.97 | 4.46 |
| Quality instructions | 0.945 | 4.34 | 0.61 | 0.14 | 4.10 | 4.58 |
| Culture of learning and professional behaviors | 0.951 | 4.34 | 0.62 | 0.14 | 4.11 | 4.57 |
| Connections to external communities | 0.961 | 4.17 | 0.74 | 0.15 | 3.93 | 4.41 |
| Systemic performance accountability | 0.957 | 4.31 | 0.66 | 0.14 | 4.08 | 4.53 |
| Planning | 0.942 | 4.33 | 0.58 | 0.14 | 4.11 | 4.56 |
| Implementing | 0.941 | 4.29 | 0.64 | 0.15 | 4.03 | 4.55 |
| Supporting | 0.935 | 4.29 | 0.63 | 0.16 | 4.03 | 4.56 |
| Advocating | 0.934 | 4.28 | 0.63 | 0.16 | 4.01 | 4.55 |
| Communicating | 0.949 | 4.30 | 0.65 | 0.15 | 4.06 | 4.54 |
| Monitoring | 0.949 | 4.22 | 0.66 | 0.15 | 3.97 | 4.47 |
| Full scale | 0.957 | 4.35 | 0.58 | 0.12 | 4.15 | 4.55 |

**Table VI.**
VAL-ED scale reliability

Standard errors of measurement were calculated for the mean scores of components, processes, and the full-scale average. The standard errors were very low and the confidence intervals for the mean scores had relatively small ranges, providing further evidence for strong internal reliability and the accuracy of the assessment results.

The inter-rater correlation results are shown in Table VII. Inter-rater reliability measures show how consistently various groups rate the same person using the assessment instrument. School-level correlations among three sets of VAL-ED scale rating results – average teacher rating for the principal, the principal self-rating, and the rating of the supervisor on the principal – were obtained from the primary sample used for the factor analysis. Correlations ranging between 0 and 0.29 were considered low; correlations $\geqslant 0.30$ to 0.59 were considered moderate; and correlations $> 0.60$ were considered high. Correlations among the ratings of three parties, teachers, principals,

| | Teacher - principal | Teacher-supervisor | Principal-supervisor | Teacher-teacher |
|---|---|---|---|---|
| *Core components* | | | | |
| High standards | 0.13 | 0.22 | 0.37 | 0.07 |
| Rigorous curriculum | 0.00 | −0.03 | 0.07 | 0.40 |
| Quality instruction | −0.02 | 0.06 | 0.39 | 0.20 |
| Culture of learning | 0.13 | −0.05 | −0.08 | 0.30 |
| External communities | −0.06 | 0.07 | −0.06 | 0.40 |
| Performance accountability | −0.01 | 0.27 | −0.19 | 0.05 |
| *Key processes* | | | | |
| Planning | 0.05 | 0.15 | 0.12 | 0.58 |
| Implementing | −0.13 | −0.04 | 0.06 | 0.18 |
| Supporting | −0.03 | 0.06 | −0.12 | 0.55 |
| Advocating | 0.17 | −0.12 | 0.09 | 0.29 |
| Communicating | 0.06 | 0.23 | 0.33 | 0.31 |
| Monitoring | −0.06 | 0.10 | −0.01 | 0.59 |
| Total mean score | 0.15 | 0.00 | 0.04 | 0.37 |

**Note:** $n = 18$

**Table VII.**
VAL-ED inter-rater school-level mean score correlations

المنارة للاستشارات
www.mana

and supervisors, were obtained for the 18 schools. The unit of analysis is the school-level mean score on the principal. The correlation was 0.15 for the total mean score, and below 0.15 for all the components and processes. Correlations were even slightly negative for three components and three processes. Correlations between teacher and supervisor scores were similar to those between teachers and principals in that the mean score correlation was minimal (0.00), and there were some scales with negative but very small correlations. As for correlations between principal and supervisor scores, the mean score correlation was also at a low 0.04, and there were three components and two processes with small, negative correlations. However, three correlations were higher than 0.30: high standards for student learning (0.37), quality instruction (0.39), and communicating (0.33) – The principals and supervisors appeared to have more agreement on their ratings in these three areas. In addition, a random pair of teachers at each of the 18 schools was selected to measure the teacher-teacher rating correlation. This set of data examined the inter-rater reliability of the assessment when it was used by respondents in the same role. The results are presented in the last column of Table VII. The correlations between two randomly selected teachers in each school rating the same principal were much higher than the teacher-principal, principal-supervisor, and teacher-supervisor correlations. The correlation of the total mean score was 0.37. A majority of the sub-scale correlations for the components and processes were in the moderate level of between 0.30 and 0.59. These inter-rater reliability results indicated that teachers, principals, and supervisors tend to rate the effectiveness of the principal differently, and teachers within the same schools tended to have more agreement on how the principal performed in their school.

The finding of low inter-rater correlations coincides with the results from the pilot testing and field study of the VAL-ED in the USA (Porter *et al.*, 2010a, b). Psychology and management research report a similar picture of low inter-rater correlation among respondent groups (Atwater *et al.*, 1998; Harris and Schaubroeck, 1988) as the results can be attributed to not simply measurement error but potentially more substantive explanations: "(*a*) systematic differences in what is observed, (*b*) systematic differences in access to information other than observations of performance, (*c*) systematic differences in expertise in interpreting what is observed, and (*d*) systematic differences in evaluating what is observed" (Murphy and Deshon, 2000, p. 822). We argue that this is precisely the added value of "360-degree" evaluations where multiple and non-redundant viewpoints can be captured.

We also found that the level of discrepancies among supervisors, principals, and teachers regarding the core components and key processes varied by school. In other words, the variations among the three parties were different depending on which principal was assessed. Data showed that principals from the 18 schools (except for the one school that did not have principal and supervisor ratings) received very different ratings from their supervisors and teachers. They also rated themselves very differently from one and other.

The between-school variation of the scores showed that the assessment instrument was able to pick up possible differences in terms of leadership effectiveness among the principals of the participating schools, a very encouraging finding about the instrument itself. The within-school variation among the three parties assessing the same principal and the different patterns exhibited in different schools pointed to possible associations between school characteristics and leadership assessment results. Other factors might have been at play, school type, principal experience, and

المنارة للاستشارات

www.mana

student characteristics, to name a few. Despite the complexity, the results demonstrate that the inter-rater correlations of the assessment scores have to be examined carefully with awareness and consideration of many other factors. A low inter-rater correlation of the overall ratings among principals, supervisors, and teachers did not necessary diminish the reliability of the instrument until we looked at how the rating consistency vary in different settings.

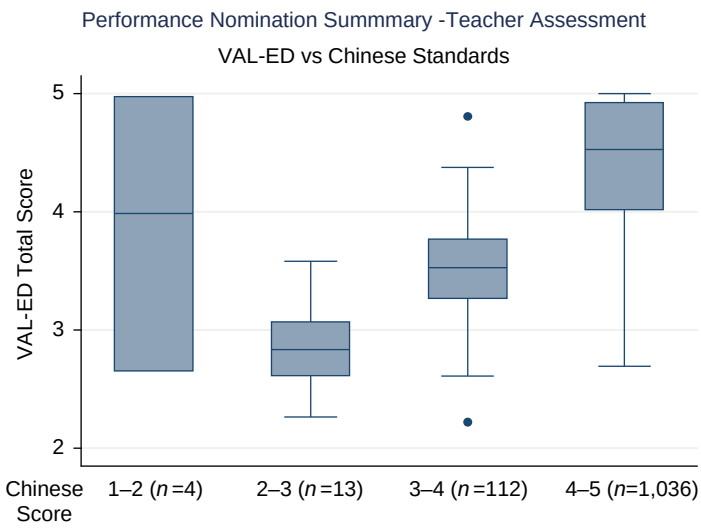
*Criterion validity.* In psychometrics, criterion validity is a measure of how well a new assessment could predict an outcome measured by an existing assessment. Criterion validity can be depicted by concurrent and predictive validity. In this case, the validity of the VAL-ED items was gauged by comparing their results to the Chinese criteria, assessed simultaneously. Specifically, we compared the results of the VAL-ED to the judgment of principal effectiveness using a set of domestic criteria by the respondents. Principals, teachers, and supervisors were asked to rate the assessed principal on the four dimensions of school leadership performance, values, abilities, diligence, and achievement, which were widely used as broad categories of principal evaluation criteria in China (MOE, 1992; Zhao and Wang, 2007). The rating scale was set on the same Likert response levels used in the VAL-ED, from 1 to 5, for equivalence comparison. The four dimensions were listed as four separate items after the 72 VAL-ED items (see Appendix 2.3). The teacher, principal, and supervisor samples for this study were the same as for the factor analytic study.

The dispersion of mean scores using both sets of criteria – the VAL-ED and the four-dimension Chinese standards – were analyzed. We used box-and whisker plots to demonstrate the extent to which the construct domain covered by the LCL framework coincided with how effective leadership is perceived in the Chinese setting. Specifically, if the two sets of criteria converged perfectly, the box-and-whisker quartile ranges of the VAL-ED effectiveness total scores (average of the 72 items) should reflect the continuum of the performance categories for the four-dimension Chinese standards.

Figure 3 illustrates the convergence of the VAL-ED rating criteria and the Chinese criteria. The mean scores for the four-item Chinese standards were grouped into four levels ($<2.0$ for Level 1, $\geqslant 2$ and $<3$ for Level 2, $\geqslant 3$ and $<4$ for Level 3, $\geqslant 4$ and $\leqslant 5$ for Level 4). VAL-ED mean scores were sorted by these levels and plotted.

The results are more interesting with the teacher assessment scores with a large number of observations ($n = 1{,}165$) and demonstrated variation (Figure 3). Overall, the VAL-ED ratings were highly skewed to the right with most of the scores above average, and very few ratings fell below 3.0. The first box plot had only four observations with the average Chinese standard scores at below 2.0 and the VAL-ED scores were spread out between 2.6 and 5. The same skewing occurred with the Chinese standard scores. The other three box plots for the VAL-ED scores were nicely aligned in an upward pattern, reflecting the concurrent increase of the Chinese standard scores. The largest group belonged to Level 4 where the Chinese standard scores are between 4 and 5 with 1,036 observations. In this group, observations between the 75th percentile (upper hinge) and the 25th percentile (lower hinge) took practically the full range between 4 and 5 on the VAL-ED score line, and the median line was at 4.52, almost a perfect midpoint between 4 and 5. Observations below the 25th percentile, however, included VAL-ED scores ranging from 4.01 to 2.69, indicating that the VAL-ED scores covered a wider range of variation in teachers' perception of their principals' effectiveness. The same could be said about the two box plots for Level 2 and Level 3, where the Chinese standard ratings matched the box positions of

المنارة للاستشارات
www.mana




**Note:** *n*=1,165

**Figure 3.**
Criterion validity check using the Chinese standards – teacher assessment

VAL-ED scores representing observations above the 25th percentile and below the 75th percentile, meanwhile the whiskers of the boxes illustrated wider ranges of the variation in the VAL-ED scores than those for the Chinese standards.

The results show that principals who received high scores based on the VAL-ED scale tended to receive comparable high scores based on the Chinese standards, and vice versa. The convergence of the VAL-ED mean scores and the mean sores for the Chinese standards provided evidence for the criterion validity of the VAL-ED as an instrument that assessed principal effectiveness in the Chinese urban school setting.

## Discussion

The extent and the depth of this study were limited, partially due to time and resource constraints, partially due to the exploratory nature of the work and some insights were only gained post facto. Findings of this paper provide opportunities for future research to further explore the possibility of adapting the LCL framework and the assessment instrument theoretically and empirically.

In this cross-cultural validity study, we were able to identify a set of general and common elements of school leadership using the LCL framework in the sampled Chinese urban schools. Input from the Chinese researchers, principals, school directors and teachers consistently confirmed that there was a strong alignment on the overarching goal of improving student social and academic learning through setting high standards, providing rigorous curriculum and quality instruction, and enhancing the professional learning culture in schools.

Our findings also highlighted noteworthy differences between the two educational systems that are rooted in a wide range of factors such as economic and demographic conditions, governance structure, and how student achievement has been defined and measured in the recent decades. For the VAL-ED to be a useful tool of assessing principal effectiveness for student learning, the LCL framework and the VAL-ED items will need to be modified. While the two-dimension conception of leadership was well

المنارة للاستشارات

www.mana

received by the Chinese participants in the study, some aspects of the core components and key processes could become conceptually clearer and the instrument items more robust if we take national and local contextual factors into consideration.

The finding of no complete cross-cultural fit of the VAL-ED theoretical framework and assessment without further modification is expected. The pendulum swings of educational reform efforts in both countries are striving to achieve the common objective of improved student learning. Educational policy makers, researchers and practitioners can benefit from learning the past and current practices of both countries and draw useful lessons from the information. In the USA, the reauthorization of NCLB has faced criticisms of overly restricting the school curriculum to focus only on the tested subjects and driving out of creative subjects such as art and literature (Burroughs *et al.*, 2005; Rothstein, 2004). In China, the push for ouality-oriented education and curriculum reform has met mounting difficulties in measuring intangible successes in student character-building, decreased efficiency of classroom activities, and widening gap between schools of high- and low-teacher capacities (Cravens *et al.*, 2011).

Well grounded in school leadership literature, the validity and reliability of the VAL-ED assessment are still going through large-scale and longitudinal studies so that more evidence can be obtained on the predictive power of the instrument for effective principals and successful schools (Chu and Cravens, 2012). Should there be an adapted theoretical framework for the Chinese setting, it may align with the US version on the level of broad definitions of leadership and reflect the common goals of education, but it must capture the deeper differences in the "directional focus" of the nation's reform imperative. That is, the major educational reform efforts taking place in China that emphasize social learning, creative thinking, and physical health must be reflected in the content domain of the assessment for it to be meaningful and useful.

Building a meaningful and practical knowledge base for Chinese school principals will be a long and challenging journey. There have been calls for caution and criticism from leading researchers on wholesale importation and implementation of educational theories without considering local context (Chu, 2003; Gao *et al.*, 2006; Zhe, 2006; Zhe and Li, 2006). More importantly, the realization of new leadership practices such as promoting creativity and connecting with external communities will not be possible without the condition of a changed accountability scheme and systemic support. Many of the Chinese educational reform mandates are yet to be implemented and cannot be fully achieved without significant changes to the current national examination system (Chu and Cravens, 2012). As long as the citywide and nationwide entrance examination system is intact, much of the educational reform initiatives will be difficult for schools and principals to implement. Furthermore, performance goals established based on the new professional standards that will be impossible for principals to reach, thus making the assessment based on such standards meaningless even potentially harmful to improving student learning. In other words, the adoption and modification of the leadership framework and assessment cannot happen without a relatively stable policy environment with well-established reform objectives that are backed with systemic support.

This paper is as much about the actual fit of the LCL framework and the VAL-ED as about seeking suitable pathways to compare and possibly transfer educational theories and models cross-culturally. We have laid out a set of analytic strategies that employ the rigor of conventional psychometric measures for instrument validation, using both qualitative and quantitative approaches, to address the complexity of defining and evaluating leadership cross-culturally.

المنارة للاستشارات
www.mana

Hard questions that pinpoint the issues at the center of a transforming nation must be asked: What are the real standards that a leadership assessment instrument should be aligned with to achieve construct validity, the current practice or the policy intent? What if there are gaps between the new standards by which principals are held accountable and the reality that principals are in? Is valid and reliable leadership assessment possible in a changing policy environment? If the standards set by the leadership assessment are to drive the behaviors of school leaders toward reaching the objectives, how should the "reality gap" be taken into consideration? How should the results of such assessment be used, for evaluations, for professional development, or both?

We hope that the results of this paper will provide some insight on the topics of cross-cultural learning in school leadership, specifically, addressing the questions of if and how leadership frameworks and their applications may be compared and understood. By examining whether and to what extent the dimensions of the LCL framework represent a more fundamental aspect of the educational experience instead of being idiosyncratic of one cultural setting, the findings suggest that the null hypothesis of complete construct equivalence are to be rejected, and culture-specific differences do exist. However, despite the differences due to socio-cultural reasons, the findings also confirm that there are significant elements of the leadership domain that are shared cross-culturally, serving as the important condition of cross-cultural learning exchange. The attention to the alignment of the theoretical framework and the validity and reliability of the assessment instrument provides a baseline for the cross-cultural comparison, and for sharing experiences and practices in areas such as measuring the effectiveness of school leadership. This is where the dual purposes of the paper converge and become meaningful.

## References

Astiz, M.F., Wiseman, A.W. and Baker, D. (2002), "Slouching towards decentralization: consequences of globalization for curricular control of national education systems", *Comparative Education Review*, Vol. 46 No. 1, pp. 66-88.

Atwater, L.E., Ostroff, C., Yammarino, F.J. and Fleenor, J.W. (1998), "Self-other agreement: does it really matter?", *Personnel Psychology*, Vol. 51 No. 3, pp. 577-598.

Ball, S. (1998), "Big policies/small world: an introduction to international perspectives in education policy", *Comparative Education*, Vol. 34 No. 2, pp. 119-130.

Brown, P. and Lauder, H. (1996), "Education, globalisation and economic development", *Journal of Education Policy*, Vol. 11 No. 1, pp. 1-25.

Burroughs, S., Groce, E. and Webeck, M.L. (2005), "Social studies education in the age of testing and accountability", *Educational Measurement: Issues and Practice*, Vol. 24 No. 3, pp. 13-20.

Chen, X. (2004), "Diversification: the changing trend for the role of school principal", *China Education Daily*, April 27, p. A-2.

Chen, Z. (2006), "Distributed leadership and school management transformation", *Overseas Collection*, Vol. 69 No. 4, pp. 1-10.

Cheng, Y.C. (2002), "Educational reforms in the Asia-Pacific region: trends and implications for research", paper presented at the International Symposium on Globalization and Educational Governance Change in East Asia, Hong Kong, June 28.

Chu, H. (2003), "Principal professionalization: knowledge base and support systems", *Educational Administration Review*, Vol. 1 No. 1, pp. 223-252.

Chu, H. (2008), "Education equity and efficiency: the dual goals of education reform", *Educational Research*, Vol. 341 No. 6, pp. 7-13.

المنارة للاستشارات

www.mana

Chu, H. and Cravens, X.C. (2012), "Principal professional development in China: challenges, opportunities, and strategies", *Peabody Journal of Education*, Vol. 87 No. 2, pp. 178-199.

Cravens, X.C. (2012), "The viability of measuring learning-centered principal leadership in Chinese schools", *Journal of Nanjing Normal University* Vol. 181 No. 3, pp. 84-94.

Cravens, X.C., Chu, H. and Zhao, Q. (2011), "Defining school effectiveness in the reform for quality-oriented education", in Huang, T. and Wiseman, A.W. (Eds), *The Impact and Transformation of Education Policy in China (International Perspectives on Education and Society Series)*, Emerald Publishing, Bingley, Vol. 15, pp. 153-185.

Cronbach, L.J. (1951), "Coefficient alpha and the internal structure of tests", *Psychometrika*, Vol. 16 No. 3, pp. 297-334.

Cronbach, L.J. and Meehl, P.E. (1955), "Construct validity in psychological tests", *Psychological Bulletin*, Vol. 52 No. 4, pp. 281-302.

Desimone, L. (2002), "How can comprehensive school reform models be successfully implemented?", *Review of Educational Research*, Vol. 72 No. 3, pp. 433-479.

Dimmock, C. and Walker, A. (2000), "Developing comparative and international educational leadership and management: a cross-cultural model", *School Leadership and Management*, Vol. 20 No. 2, pp. 143-160.

Dorfman, P.W. and Howell, J.P. (1997), "Leadership in western and Asian countries: commonalities and differences in effective leadership processes across cultures", *Leadership Quarterly*, Vol. 8 No. 3, pp. 233-274.

Du, P. (2004), "Highly effective principal leadership as the important condition for curriculum reform", *Research on Educational Development*, Vol. 11 No. 1, pp. 28-31.

Elliott, S.N. (2008), "VAL-ED items and scores: types, interpretations, and error", paper presented at the Annual Meeting of American Educational Research Association New York, NY, March 20.

Gao, R., Wang, B. and Lin, W. (2006), "The foundation of moral theory – research and reflection on leadership theory in China", *Science and Technology Management Research*, Vol. 6 No. 1, pp. 142-147.

Goldring, E., Cravens, X.C., Murphy, J., Elliott, S.N., Carson, B. and Porter, A.C. (2009a), "The evaluation of principals: what and how do states and districts assess leadership?", *The Elementary School Journal*, Vol. 110 No. 1, pp. 29-39.

Goldring, E., Porter, A., Murphy, J., Elliott, S.N. and Cravens, X. (2009b), "Assessing learning-centered leadership: connections to research, professional standards, and current practices", *Leadership and Policy in Schools*, Vol. 8 No. 1, pp. 1-36.

Gu, M. (2010), "Promoting quality-oriented education: the strategic theme of educational reform", *Guangmin Daily*, February 24, pp. 1-2.

Hair, J.F., Anderson, R.E., Tatham, R.l. and Black, W.C. (1998), *Multivariate Data Analysis*, 5th ed., Prentice Hall, Upper Saddle River, NJ.

Hallinger, P. (2011), "Developing a knowledge base for educational leadership and management in East Asia", *School Leadership and Management*, Vol. 31 No. 4, pp. 305-320.

Hallinger, P. and Kantamara, P. (2001), "Learning to lead global changes in local cultures: designing a computer-based simulation for Thai school leaders", *Journal of Educational Administration*, Vol. 39 No. 3, pp. 197-220.

Hallinger, P. and Leithwood, K. (1998), "Unseen forces: the impact of social culture on school leadership", *Peabody Journal of Education*, Vol. 73 No. 2, pp. 126-151.

Hallinger, P. and Murphy, J. (1987), "Instructional leadership in the school context", in Greenfiled, W. (Ed.), *Instructional Leadership: Concepts, Issues, and Controversies*, Allyn and Bacon, Boston, MA, pp. 179-203.

المنارة للاستشارات

www.mana

Hallinger, P., Walker, A. and Bajunid, I. (2005), "Educational leadership in East Asia: implications for education in a global society", *UCEA Review*, Vol. 1 No. 1, pp. 1-4.

Hambleton, R. (2002), "Adapting achievement tests into multiple languages for international assessments", in Porter, A.C. and Gamoran, A. (Eds), *Methodological Advances in Cross-National Surveys of Educational Achievement*, National Academy Press, Washington, DC, pp. 58-79.

Hambleton, R., Merenda, P. and Spielberger, C. (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, Lawrence Erlbaum Associates, Mahwah, NJ.

Harris, M.M. and Schaubroeck, J. (1988), "A meta-analysis of self-supervisor, self-peer, and peer supervisor ratings", *Personnel Psychology*, Vol. 41 No. 1, pp. 43-62.

Heck, R.H. (1996), "Leadership and culture: conceptual and methodological issues in comparing models across cultural settings", *Journal of Educational Administration*, Vol. 34 No. 5, pp. 74-97.

Heck, R.H., Larsen, T.J. and Marcoulides, G.A. (1990), "Principal leadership and school achievement: validation of a causal model", *Educational Administration Quarterly*, Vol. 26 No. 2, pp. 94-125.

Heck, R.H. and Marcoulides, G.A. (1996), "The assessment of principal performance: a multilevel evaluation approach", *Journal of Personnel Evaluation in Education*, Vol. 10 No. 1, pp. 11-28.

Hirosato, Y. (2001), "New challenges for educational development and cooperation in Asia in the 21st century: building indigenous capacity for education reforms", *Journal of International Cooperation in Education*, Vol. 4 No. 2, pp. 1-20.

Huang, W. (2004), "Principal: from the role of an administrative position to the role of a management profession", *Educational Administration Review*, Vol. 2 No. 1, pp. 289-300.

Kaiser, H.F. (1958), "The varimax criterion for analytic rotation in factor analysis", *Psychometrics*, Vol. 23 No. 3, pp. 187-201.

Lee, V.E. and Bryk, A.S. (1989), "A multilevel model of the social distribution of highschool achievement", *Sociology of Education*, Vol. 62 No. 3, pp. 172-192.

Leithwood, K. and Jantzi, D. (1993), "Transformational leadership: how principals can help reform school cultures", *School Effectiveness and School Improvement*, Vol. 1 No. 4, pp. 249-280.

Li, Y. (2004), "Reorientation for principals' role and function", *Educational Administration Review*, Vol. 2 No. 1, pp. 301-312.

Linn, R.L. (2000), "Assessment and accountability", *Educational Researcher*, Vol. 29 No. 2, pp. 4-16.

Maxwell, L.A. (2009), "Review backs new tool for principal evaluation: Vanderbilt's new assessment receives high marks, but many are not linked to students' learning", *Education Week*, Vol. 29 No. 16, p. 8.

Ministry of Education (MOE) (1992), *The Draft Opinions on Enhancing the Development of School Principals in the Nation*, MOE, Beijing.

Murphy, J., Elliott, S.N., Goldring, E. and Porter, A.C. (2007), "Leadership for learning: a research-based model and taxonomy of behaviors", *School Leadership and Management*, Vol. 27 No. 2, pp. 179-201.

Murphy, J.F., Goldring, E.B., Cravens, X.C., Elliott, S.N. and Porter, A.C. (2011), "The Vanderbilt assessment of leadership in education: measuring learning-centered leadership", *Journal of East China Normal University*, Vol. 29 No. 1, pp. 1-10.

Murphy, K.R. and Deshon, R. (2000), "Interrater correlations do not estimate the reliability of job performance ratings", *Personnel Psychology*, Vol. 53 No. 4, pp. 873-900.

المنارة للاستشارات

www.mana

National Center for Education Development Research (NCEDR) (2008), *2007 Green Paper on Education in China*, Educational Science Publishing House, Beijing.

National Governors Association (2008), *Benchmarking for success: Ensuring US Students Receive a World-Class Education*, National Governors Association, Washington, DC.

Ng, C.-H. (2006), "Exploring the linkage between reforms and learning in the Asia-Pacific Region: an activity theory perspective", in Ng, C.-H.C. and Renshaw, P.D. (Eds.), *Education in the Asia-Pacific Region: Issues, Concerns and Prospects: Reforming Learning*, Springer, Dordrecht, pp. 23-40.

Peterson, P.E., Woessmann, L., Hanushek, E.A. and Lastra-Anadón, C.X. (2011), *Globally Challenges: Are American Students Ready to Compete?*, Taubman Center for State and Local Government, Harvard Business School, Boston, MA.

Porter, A., Goldring, E., Murphy, J., Elliott, S.N. and Cravens, X. (2006), *A Conceptual Framework for the Assessment of Principal and Team School Leadership*, Wallace Foundation, New York, NY.

Porter, A.C., Polikoff, M., Goldring, E.B., Murphy, J., Elliott, S.N. and May, H. (2010a), "Developing a psychometrically sound assessment of school leadership: the VAL-ED as a case study", *Educational Administration Quarterly*, Vol. 46 No. 2, pp. 135-173.

Porter, A.C., Polikoff, M.S., Goldring, E., Murphy, J., Elliott, S.N. and May, H. (2010b), "Investigating the validity and reliability of the Vanderbilt assessment of leadership in education", *Elementary School Journal*, Vol. 111 No. 2, pp. 282-313.

Portin, B.S., Feldman, S. and Knapp, M.S. (2006), *Purposes, Uses, and Practices of Leadership Assessment in Education*, Wallace Foundation, New York, NY.

Rothstein, R. (2004), *Class and Schools: Using Social, Economic, and Educational Reform to Close the Achievement Gap*, Economic Policy Institute, Washington, DC.

Sergiovanni, T.J. (1992), *Moral Leadership*, Jossey-Bass, San Francisco, CA.

Spillane, J.P. (2006), *Distributed Leadership*, Jossey-Bass, San Francisco, CA.

State Council of the Chinese Central Government (1993), *Guidelines for the Reform and Development of Education in China*, State Council of the Chinese Central Government, Beijing.

State Council of the Chinese Central Government (1999), *The Decisions on Deepening Educational Reform and Promoting Quality-Orientated Education*, State Council of the Chinese Central Government, Beijing.

Stevens, J. (1996), *Applied Multivariate Statistics for the Social Sciences*, 3rd ed., Lawrence Erlbaum, Mahwah, NJ.

Thompson, B. (2004), *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*, American Psychological Association, Washington, DC.

Walker, A., Hu, R. and Qian, H. (2012), "Principal leadership in China: an initial review", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, Vol. 23 No. 4, pp. 369-399.

Westbury, H. (1992), "Comparing American and Japanese achievement: is the United States really a low achiever?", *Education Researcher*, Vol. 21 No. 5, pp. 18-24.

van Widenfelt, B.M., Treffers, P.D.A., de Beurs, E., Siebelink, B.M. and Koudijs, E. (2005), "Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families", *Clinical Child and Family Psychology Review*, Vol. 8 No. 2, pp. 135-148.

Zhao, Q. and Wang, M. (2007), "Construction of performance appraisal system for primary and secondary school principals-a case study in Beijing", *Educational Administration Review*, Vol. 4 No. 1, pp. 194-244.

المنارة للاستشارات

www.mana

Zhe, Y. (2006), "Review of contemporary western school leadership theories", *Journal of Wenzhou Normal University*, Vol. 25 No. 3, pp. 1-12.

Zhe, Y. and Li, C. (2006), "Comparative theory and practice of school leadership management model", *Journal of Wenzhou Normal University*, Vol. 27 No. 1, pp. 113-116.

Zhou, J. (2004), "The full implementation of the 2003-2007 action plan for invigorating education", available at: www.edu.cn/20040324/3102182.shtml (accessed May 8, 2008).

## Appendix 1. LCL framework

Definitions of core components and key processes (Porter *et al.*, 2006)

*Core components of school performance*

The first dimension consists of core components of school performance and has the following elements:

- High standards for student performance – there are individual, team and school goals for rigoros student academic and social learning.
- Rigoros curriculum (content) – there is ambitious academic content provided to all students in core academic subjects.
- Quality instruction (pedagogy) – there are effective instructional practices that maximize student academic and social learning.
- Culture of learning and professional behavior – there are integrated communities of professional practice in the service of student academic and social learning. There is a healthy school environment in which student learning is the central focus.
- Connections to external communities – there are linkages to people and institutions in the community that advance academic and social learning.
- Systemic performance accountability – leadership holds self and others responsible for realizing high standards of performance for student academic and social learning. There is individual and collective responsibility among the professional staff and students.

*Key processes of leadership*

The second dimension defines the leadership behaviors that can lead to producing each core component of school performance. These key processes are:

- Planning – articulate shared direction and coherent policies, practices and procedures for realizing high standards of student performance.
- Implementing – engage people, ideas and resources to put into practice the activities necessary to realize high standards for student performance.
- Supporting – create enabling conditions; secure and use the financial, political, technological, human and social capital necessary to promote academic and social learning.
- Advocating – act on behalf of the diverse needs of students within and beyond the school.
- Communicating – develop, utilize and maintain systems of exchange among members of the school and with its external communities.
- Monitoring – systematically collect and analyze data to make judgments that guide decisions and actions for continuous improvement.

المنارة للاستشارات

www.mana

**Appendix 2.**
*2.1: Instructions for the VAL-ED forms used in studies 1-3 (the originals are in Chinese)*

Dear teacher (or principal, or supervisor):

You are invited to participate in this survey as a part of the comparative study on school leadership assessment in the United States and China. We want to find out the fit and feasibility of a principal evaluation system in urban Chinese schools. This evaluation system uses assessment results from principal self-evaluation, teachers' evaluation of the principal, and the evaluation from the principal's supervisor. The Vanderbilt Assessment of Leadership in Education (VAL-ED) measures the effectiveness of a principal's key leadership behaviors that influence teacher performance and student learning. You will be asked to make effectiveness ratings for each of 72 leadership behaviors based on evidence from the current school year. As stated in the consent form, results of the assessment will only be used for research purposes, not for any official evaluation of the school and the principal."

1. Read each item describing a leadership behavior. In some cases, the principal may not have actually performed the behavior, but he or she has ensured that it was done by others in the school. Either way the behavior should be rated.

2. Check (√) the key **Sources of Evidence** you use for the basis of your assessment. Note, at least one source of evidence must be checked for an item before you make an Effectiveness rating. If you check **No Evidence**, then **Ineffective** or **Don't Know** must be marked in the Effectiveness column.

3. If you check **any sources of evidence other than No Evidence**, always make an effectiveness rating even if you must **estimate** the effectiveness of the behavior. The number of **Sources of Evidence** checked is **not** indicative of the effectiveness rating.

4. Mark one **Effectiveness Rating** circle to indicate how effectively the behavior was performed.

   **Outstandingly effective** means the principal (or the principal's designee) has carried out a particular behavior (e.g., providing necessary support) with a very strong, positive effect on the targeted area of school activity (e.g., rigorous curriculum).

   **Ineffective** means the principal (or the principal's designee) has either not done the particular behavior (e.g., not provided necessary support) or has carried out the behavior with very low quality that does not have a positive effect on the targeted area of school activity (e.g., rigorous curriculum).

(For teachers)

*Years as teacher:* _____ *Years at this school :* _____

*Gender :* _____

*Subject taught:*

*1. Math or Science* _____ *2.Chinese or English* _ _ _ _ _ _ _

*3. Politics* _____ *4.Hisotry or Geography* _____

*5. Other* _ _ _ _ _ _ _

(For principals)

*Years as teacher:* _____ *Years as School Administrator :* _____

*Years as assistant principal :* _____ *Years as principal:* _____

*Years as principal in this school:* _____ *Gender:* _____

www.mana

Rating Example:

| | | Effectiveness Rating (Mark One Circle to Indicate How effective) | | | | | | Sources of Evidence (You may check more than one) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Quality Instruction | Ineffective | Minimally Effective | Satisfactorily Effective | Highly Effective | Outstandingly Effective | Don't Know | Report from others | Personal observations | School documents | School projects or activities | Other Sources | No evidence |
| | *How effective is the principal at ensuring the school . . .* | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| Supporting | 29. supports collaboration among faculty to improve instruction that maximizes student learning. | | | | | | | | | | | | |
| | 30. supports teachers' opportunities to improve their instructional practices. | | | | | | | | | | | | |

*2.2: The 72-item VAL-ED scale*

| Item no. | How effective is the principal at ensuring the school … |
|---|---|
| 1 | Plans rigorous growth targets in learning for all students |
| 2 | Plans targets of faculty performance that emphasize improvement in student learning |
| 3 | Creates buy-in among faculty for actions required to promote high standards of learning |
| 4 | Creates expectations that faculty maintain high standards for student learning |
| 5 | Encourages students to successfully achieve rigorous goals for student learning |
| 6 | Supports teachers in meeting school goals |
| 7 | Advocates for high standards for student learning when writing and implementing Individualized Education Plans (IEPs) |
| 8 | Challenges low expectations for students with special needs |
| 9 | Communicates rigorous goals for student learning to faculty |
| 10 | Communicates with families and the community about goals for rigorous student learning |
| 11 | Monitors student learning against high standards of achievement |
| 12 | Monitors disaggregated test results |
| 13 | Develops a rigorous curriculum for all students |
| 14 | Plans access to rigorous curricula for students with special needs |
| 15 | Creates rigorous sequences of learning experiences/courses |
| 16 | Implements a rigorous curriculum in all classes |
| 17 | Secures the teaching materials necessary for a rigorous curriculum |
| 18 | Supports teachers to teach a curriculum consistent with state and national content standards |
| 19 | Advocates a rigorous curriculum that honors the diversity of students and their families |
| 20 | Challenges faculty to teach a rigorous curriculum to students at risk of failure |
| 21 | Discusses state curriculum frameworks |
| 22 | Discusses the importance of addressing the same academic content in special and regular programs |
| 23 | Evaluates the extent to which all students complete a rigorous curricular program |
| 24 | Evaluates the rigor of the curriculum |
| 25 | Plans instructional services for students with special needs using assessment data |
| 26 | Plans a schedule that enables quality instruction |

المنارة للاستشارات

www.mana

27 Coordinates efforts to improve instruction in all classes
28 Recruits teachers with the expertise to deliver instruction that maximizes student learning
29 Supports collaboration among faculty to improve instruction that maximizes student learning
30 Supports teachers' opportunities to improve their instructional practices
31 Advocates for all students to regularly experience effective instruction
32 Advocates opportunities for high quality instruction beyond the regular school day and school year
33 Discusses instructional practices during faculty meetings
34 Communicates with faculty about removing barriers that prevent students from experiencing quality instruction
35 Evaluates how instructional time is used
36 Evaluates teachers' instructional practices
37 Plans programs and policies that promote discipline and order
38 Plans for a positive environment in which student learning is the central focus
39 Implements a learning environment in which all students are known and cared for
40 Builds a culture that honors academic achievement
41 Allocates resources to build a culture focused on student learning
42 Supports collaborative teams to improve instruction
43 Advocates a culture of learning that respects diversity of students
44 Advocates for students to be involved in the school community
45 Communicates with parents about the aspects of a positive school culture
46 Discusses standards of professional behavior with faculty
47 Monitors the participation of every student in social and academic activities
48 Assesses the culture of the school from students' perspectives
49 Develops a plan for school/community relations that revolves around the academic mission
50 Develops a plan for community outreach programs consistent with instructional goals
51 Implements programs to help address community needs
52 Builds business partnerships to support social and academic learning
53 Secures additional resources through partnering with external agencies to enhance teaching and learning
54 Allocates resources that build family and community partnerships to advance student learning
55 Promotes mechanisms for reaching families who are least comfortable at school
56 Challenges teachers to work with community agencies to support students with low achievement
57 Listens to feedback from the community
58 Listens to the diverse opinions and needs of all families
59 Collects information to learn about resources and assets in the community
60 Monitors the effectiveness of community-school connections
61 Develops a plan for individual and collective accountability among faculty for student learning
62 Develops a plan emphasizing accountability to stakeholders for student academic and social learning
63 Uses faculty input to create methods to hold faculty accountable
64 Implements social and academic accountability equitably for all students
65 Allocates time to evaluate student learning
66 Allocates time to evaluate faculty for student learning
67 Challenges faculty who attribute student failure to others
68 Advocates that all students are accountable for achieving high levels of performance in both academic and social learning
69 Discusses progress toward meeting school goals with parents
70 Communicates to faculty how accountability results will be used for school improvement
71 Analyzes the influence of faculty evaluations on the rigor of the curriculum
72 Monitors the accuracy and appropriateness of data used for student accountability

المنارة للاستشارات
www.mana

*2.3: The four-item scale for the Chinese standards (the original is in Chinese)*
In Chinese schools, the following four categories are often used to evaluate the performance of a principal. Please check one effectiveness rating to indicate how effectively the principal has performed in each category.

| | Ineffective 1 | Minimally effective 2 | Satisfactorily effective 3 | Highly effective 4 | Outstandingly effective 5 |
|---|---|---|---|---|---|
| Value | | | | | |
| Ability | | | | | |
| Diligence | | | | | |
| Achievement | | | | | |

## Appendix 3. Expert-panel alignment rating form (the full version is in Chinese)

Part I. Relevance and Importance of the 72 VAL-ED Leadership Behavior Items
(Items Not Listed)

| VAL-ED Items：(Learning-Centered Leadership Behaviors) | Please rate the relevance of this leadership behavior to the current practice of Chinese urban school principals (the extent to which it is reality) | | | | | Your comments on the item, and your suggestions for better translation |
|---|---|---|---|---|---|---|
| | Little to None | A Little | Some what | Much | Vern Much | |
| | 1 | 2 | 3 | 4 | 5 | |

| VAL-ED Items：(Learning-Centered Leadership Behaviors) | Please rate the importance of this leadership behavior to the success of Chinese urban schools and students (the extent to which it should be practiced) | | | | | Your comments on the item, and your suggestions for better translation |
|---|---|---|---|---|---|---|
| | Not Important | Not Very Important | Some what | Quite Important | Very Important | |
| | 1 | 2 | 3 | 4 | 5 | |

Part II: Suggestions for Core Components and Key Processes

In your opinion, are there any core components or key processes of learning-centered leadership that might be missing from the VAL-ED framework if it is used in the Chinese educational context?

| Core Components | Key Processes |
|---|---|
| High standards for student performance | Planning |
| Rigorous curriculum (content) | Implementing |
| Quality instruction (pedagogy) | Supporting |
| Culture of learning and professional behavior | Advocating |
| Connections to external communities | Communicating |
| Systemic performance accountability | Monitoring |
| Your suggestions for additional components: | Your suggestions for additional processes: |

المنارة للاستشارات

www.mana

**About the author**
Dr Xiu Chen Cravens is an Assistant Professor of the practice in education policy at the Department of Leadership, Policy, and Organization and the Associate Dean for International Affair at the Peabody College of Education and Human Development, Vanderbilt University. She also serves as the China Liaison for Vanderbilt University. Dr Xiu Chen Cravens can be contacted at: xiu.cravens@vanderbilt.edu

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**

المنارة للاستشارات

www.mana

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

www.manaraa.com