THE CROSS-CULTURAL FIT

OF THE LEARNING-CENTERED LEADERSHIP FRAMEWORK

AND ASSESSMENT FOR CHINESE PRINCIPALS

By

Xiu Chen Cravens

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

August, 2008

Nashville, Tennessee

Approved:

Professor Ellen Goldring

Professor Steve Elliott

Professor Joseph Murphy

Professor James Guthrie

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Chapter   Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The stakes are high for having effective school leaders in the United States and

many countries in the world. The belief that there are general and common elements in

contemporary international educational policy (Ball, 1998; Brown & Lauder, 1996) has

also 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. The challenge, however, is to understand the complexity of cross-cultural

translation of educational theories and applications and furthermore, to get beyond the

recognition of such challenge by mapping out possible solutions to sustainable and

meaningful cross-cultural adaptation.

Brief Background of the Chinese Principal Professionalization Movement

China's impressive economic performance over the past 20 years is well known.

Less well-known are the significant changes that have occurred in education. On the

positive side, in a relatively short period of time, China has made great strides toward

eliminating illiteracy, providing nine years of basic education to up to 94% of the school-

age children for a total of about 180 million, and dramatically expanding the numbers of

students in higher education to 23 million, about 21% of high school graduates (MOE,

2005). The drivers of change in developing countries, as pointed out by Heyneman

1

(2001), include new technology, internationalization, international trade, political

changes, consolidation of industry, privatization, and demographic changes, all of which

apply to China. The massive change of the economic landscape has also brought

significant shifts in governance structure, financing schemes and accountability policies

with direct impact on the education system. In the mean time, societal mandates

pertaining to the quality of education have been elevated, where graduates of the public

school system are facing an increasingly complex labor market in need of knowledge

workers who are creative and adaptive. Consequently, the role of educational leadership,

especially the role of school principals is getting significant attention from policy makers

and educational administration scholars (Chu, 2004). A movement of principal

professionalization that aims at developing newer and stronger leadership is gaining

momentum in People's Republic of China.

To professionalize school principals in China, two main challenges exist:

providing training for preparation and development, and implementing support systems

for school leaders (Chu, 2003). For historical and political reasons, domestic theories and

empirical evidence for educational leadership are limited in China. A variety of imported

school leadership theories and models have been introduced to the Chinese research

community in the recent decades, including strands such as transformational leadership

(Leithwood, 1993), instructional leadership (Hallinger, 1987), distributed leadership

(Spillane, 2006) and moral leadership (Sergiovanni, 1992), to name a few. The challenge,

however, is to identify effective methods to learn from existing and merging leadership

theories and models for the purpose of building a meaningful and practical knowledge-

base that fits the context of Chinese schools, without wholesale importation and

implementation of educational theories (Chu, 2003; Gao, Wang & Lin, 2006; Zhe, 2004; Zhe & Li, 2006).

The "Learning-Centered Leadership" Framework: Assumptions and Predictions

This dissertation selects a theoretical framework for effective school leadership that is connected with research, standards and current practices in the United States, and explores its generalizability cross-culturally. The Learning-Centered Leadership Framework (Goldring et al., 2007, Murphy et al., 2007; Porter et al., 2006) identifies the type of leadership behaviors found in the literature on effective schools and school districts in the last three decades, and reveals two key dimensions of highly effective leadership related to student learning and achievement: *core components* and *key processes*. 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 (Conley & Goldman, 1994; Leithwood, 1994). Effective learning-centered leadership, according to this Framework, is at the intersection of the two dimensions. The Framework is the foundation of a leadership assessment instrument, the Vanderbilt Assessment of Leadership in Education™ (the VAL-ED), a tool that can be used for principal evaluation, coaching, and professional development.

Underlying the fundamental assumption that theories and models of educational leadership can be compared cross-culturally is belief that the ultimate goal of education is universal. Countries that promote the common good of public education tend to focus on student achievement, despite the inherent tension and conflicts among these values:

3

quality, efficiency, equity and choice (Coleman, 1990; Green, 1985; Wirt & Kirst, 1982). *Quality* and *efficiency* of education refer not only to what can be achieved in academic achievements, but to life and social skills of students; *equity* and *choice* of education emphasize on how well education is distributed to all students despite their diverse backgrounds. This seemingly universal assumption also extends further, to the path of realizing quality and equality of education in any cultural context. The path is believed to be through teaching and learning, both inside and outside of schools and classrooms. Furthermore, school leaders, especially school principals, play an essential role in ensuring that teaching and learning meet the needs of students and reach the established internal and external standards (Chu, 2003; Goldring et al. 2007; Murphy et al., 2007).

To maximize the benefit of learning across culturally distinctive educational settings, however, we must ask the following critical questions: Is there a universal set of effective leadership behaviors that positively influence school performance and student learning? Are there elements of leadership that are unique in each societal and cultural setting? What are some of the important societal and cultural variables that may influence the definition of effective leadership? How are predictors for effective leadership connected with important societal and cultural variables?

Dissertation Purposes and Research Questions

The purpose of this dissertation study is threefold: First, to empirically test the

theoretical fit of *the Learning-Centered Leadership Framework* (Porter et al., 2006;

Murphy et al., 2007), a conceptual framework that is aligned with professional standards

and current practices in the United States (Goldring et al., 2007), with leadership in urban

Chinese schools. Second, to examine cross-cultural evidence for the validity and

reliability of *the Vanderbilt Assessment of Leadership in Education* ™ (the VAL-ED)

scores. This instrument was developed from the *Learning-Centered Leadership*

Framework and has been validated on a sample of school leaders and teachers in the

United States. Third, to explore possible modifications to the theoretical framework and

the assessment instrument that will enhance the cross-cultural relevance and utility of the

VAL-ED.

Specific to these purposes, this dissertation asked the following research questions:

(*a*) *The Fit of the Theoretical Framework*. How well does the Learning-Centered

Leadership 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? *(b) 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? *(c) The Relevance and Utility*

of the Framework and the Instrument. Based on the results of the first two questions,

should the framework and the instrument be modified, and if yes, what may be some of

the suggestions, to enhance the cross-cultural relevance and utility of the Learning-Centered Leadership Framework and the VAL-ED?

Dissertation Overview

This dissertation is organized into five chapters. Chapter I introduces the purpose of the dissertation along with brief background information on the topic. Chapter II provides a review of literature that addresses the societal and cultural context of Chinese education, the importance of leadership assessment as an element of principal professionalization, the rationale of the Learning-Centered Leadership Framework and the VAL-ED, and the challenges of cross-cultural comparison and validation. Chapter III describes the research methods proposed in the main studies, specifically, the purpose of each study as related to the research questions, predicted results, sampling, data collection, and analysis procedures. Chapter IV reports the results of the studies that are compared with the predictions, with data tables, figures, and analysis narratives. Chapter V summarizes the results and discusses the significance and implications of the findings, limitations of the dissertation, suggestions and directions for future research, and final conclusions.

Significance of the Research

By employing both qualitative and quantitative methods to examine the validity of a particular construct to operate in a consistent way across cultures, the anticipated significance of this dissertation was twofold: First, it shed light on the extent the dimensions of Learning-Centered Leadership Framework represents a more fundamental

aspect of the educational experience instead of the idiosyncrasies of one cultural setting, providing an opportunity for creating a cohesive understanding of the leadership behaviors and processes associated with effective schools. Second, on the methodology front, this dissertation made a contribution to the body of literature that informs both researchers and practitioners on viable venues of cross-cultural learning exchange of educational theories and applications, by testing possible solutions to examining the validity and reliability of a theoretical framework and assessment in a different socio-cultural setting.

CHAPTER II

REVIEW OF LITERATURE

Understanding School Leadership in the Chinese National Context

To understand the challenges and opportunities faced by policy makers,

educational researcher and practitioners, we must first examine the societal culture and

political environment in which the Chinese education system exists and evolves. Three

aspects are key to depicting the interlocking system that fosters the construct of

leadership in the Chinese educational setting. First, core cultural beliefs, values and

ideologies that underpin organizational structures, processes and practices at the national

and local levels; second, the operationalization of leadership effectiveness in

contemporary Chinese educational settings in terms of qualifications, professional

standards, performance mandates, and job evaluations for school principals; and third,

educational policy and research efforts in China geared toward building a knowledge

base for school leadership development and establishing systemic support for school

performance improvement.

Core Cultural Beliefs, Values, and Ideologies

Cheng (1995) and Hallinger and Leithwood (1998) have argued for greater

attention to be taken of societal culture in studies of educational leadership and

educational administration. The notion of culture is defined by Dimmock (2000) as "the

enduring sets of beliefs, values and ideologies underpinning structures, processes and

practices which distinguish one group of people from another. The group of people may

be at school level (organizational culture) or at the national level (societal culture)"

(p.146).

Table 1
Comparison of Chinese and Western Cultures

Philosophical/Cultural Aspect	Western Culture	Chinese Culture*
Thinking method	Logical-analytical	Synthetic
		Dialectic
Epistemological aspect	Scientific	Metaphysical
	Knowledge-based	Inner moral
Ontological aspect	Nature	Life
Cultural aspect	Realistic	Pragmatic
	Individual ability	Community-related effort
Religious aspect	Monotheistic	Indifferentist (neutralist)

(Wong, 1998, p.108)

The word for "leader" in Chinese is depicted by two separate characters.

The first character means to lead a group with reasoning, and the other to provide

guidance. Wong (1998) provided a comparison of Chinese and Western Culture

(Greco-Roman, Judeo-Christian) as illustrated in Table 1, in which he traced the

origin of the development of early Chinese society and attempted to group these

differences according to several analytic categories. He concluded that the Chinese

are different from Westerners in many significant aspects.   For example, the early

Chinese scholars were not interested in pure abstract thinking. When they came

across abstract ideas, they always expressed them in terms of concrete or visual

objects. The ancient Chinese were also not interested in spiritual beings. This was

reflected in the works of Confucius. These habits of the Chinese to conceptualize

ideas in concrete (particular) objects and to adopt a pragmatic attitude toward life in Confucianism took on social and moral significance, which has had profound influence in Chinese societal culture. Under the influence of Confucianism, many Chinese scholars considered it their obligation to serve the state. Partly this sense of obligation grew out of necessity, following the Han dynasty (25-220 AD), the Chinese states developed a system that absorbed scholars into the civil service. After the Tang dynasty (618 -907 AD), the same sentiment of loyalty was fostered through a sophisticated system of national examinations. Subsequently and for a very long period of time, joining the civil service was the only recognized outlet of scholarly expression. Compared to the past in Europe where a man could rise to a position of high social status through a career in law, medicine, commerce, the Church, or the military, in China there was only one significant occupational hierarchy: the civil service (Fairbank, 1994). Since the New Cultural Movement in 1915 (four years later came the famous May Fourth Movement), however, many cultural elites publicly denounced the value system and the culture of the past, in particular Confucianism (Fairbank, 1994). Despite this departure from the past, cultural values change slowly and the traditional Chinese values are still influential in many Chinese societies today (Wong, 1998).

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 implications in how leaders in China think, operate, and receive their appraisals. School principals, for example, have long been considered 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 (Lin, 2003; Gao, Wang & Lin, 2006).

The Changing Economic and Policy Environment

The long-held beliefs, values and ideologies, however, are being challenged more than ever.   In education, the trend of reforms towards decentralization, marketization and diversification has become more and more important in China, particularly because China, as many other developing countries, suffers from the limitation of resources to expand its educational services to meet the diverse and increasing demands of education. In moving along this trend, some critical issues in China are emerging that challenge policy-makers, social leaders and educators to establish and implement new educational policies that are aligned with the political and economic priorities (Cheng, 2002; Chu, 2003, Chu, 2007).

Chinese education has experienced a radical transformation since 1949, with the founding of the People's Republic of China. Between 1949 and the early 1990s, traditional private education vanished in the Chinese education system. Between 1949 and the early 1980s, the collective or nationalist ideology was dominant, and individual goals were submerged in favor of social goals. During the post-1949 period, expanding educational access was consistently a focus of educational policy. Over time, the government managed to essentially achieve universal primary education, and by 1985 the major educational goal was universal compulsory education by around 2000 (People's Press, 1985). Access to higher levels of schooling was purportedly based on merit, particularly on performance on examinations. However, government spending on

education was persistently low, both in terms of national-effort and fiscal-effort indicators. This low-spending imposed a serious constraint on educational development in China and contributed partly to the changing mandates on the role of school principals in the 1990s (Tsang, 2002).

Deng Xiaoping was the supreme leader of the Central Community Party and the Chinese State from 1978 until his death in 1997. He and his followers reversed the national and educational policies of the radical faction. Through successful implementation of economic reform policies and opening-up to the outside world, the Chinese economy grew rapidly and the average living standard of the Chinese people improved substantially (Dernberger, 1999). However, reform policies since 1978 have also contributed to substantial and even widening economic disparities across areas and regions in the countries (World Bank 1998; Li & Zhao, 1999) and in education (Tsang, 1994). There are large differences in quality (and in transition rates) between a limited number of highly selective schools and regular schools. Quality differences between schools are more pronounced at the secondary level than at the primary level (Tsang, 2002).

Economic reform in China since 1978 has led to a diminished role of the State in economic production and to an increased reliance on market forces. It has been a contributing factor to educational disparities and to the necessity to seek alternative resources for the education sector. The reform was based on two financing strategies: decentralization in which financing responsibility is delegated to local governments, and diversification in which both government and non-government resources were mobilized for education. In China, resources for education are put under budgetary resources and

out-of-budget categories. Budgetary resources are from government allocation. Out-of-budget resources consist of education levies and surcharges, school fees, work study, social (domestic) contributions, and overseas contributions. Non-government resources fall under the out-of-budget category. Different localities differ significantly in their capacity to raise out-of-budget resources for education (Tsang, 1996). However, with little fiscal equalization through intergovernmental education grants, reform also brought on the expected result of substantial financial inequality among areas and regions in the country. Not only was the distribution of educational resources uneven, the total amount of national spending on education has been consistently low. Breaking the monopolistic role of the State in education was seen by some policy makers as an extension of what is happening in the economic sector (Tsang, 2002). School principals in China have the important task of raising additional revenue, for example, to augment the meager income of teachers and to improve the physical conditions of their school (Li, 2004).

By the early 1990s, universal nine-year compulsory education was accomplished in most of the urban areas in China. The focus of educational policy in urban China was shifted towards the expansion of upper-secondary education and the new emphasis on the "all-rounded" qualities of students that includes not only test scores, but citizenship, creativity and physical health (Tsang, 2002). The low-quality of some government schools, especially at the lower-secondary level, is of particular concern to both educational policy makers and to parents. There is thus motivation to link efforts to develop stronger school leadership as the key response to the transition (Chu, 2006).

"There have been two persistent tensions in the goals of educational development in post-1949 China. The first is the tension between education for promoting social

equality and education for economic efficiency; and the second is between education for inculcating socialist ideals and education for developing talents ('redness' vs. 'expertise')" (Tsang, 2002, p.16). With its support for key educational institutions, the use of the national examination for educational selection, and the emphasis on science, technology, and productive skills, the current government's policy tends to favor education for economic efficiency and for developing talents. However, there is still a strong national ideology for social equality and social goals through education. Problems in educational financing, particularly low educational spending and substantial disparity in financial health among schools and localities, remain a key issue in Chinese education today (Chu, 2007).

The Changing Role of School Principals

At the center of educational reform in China, the role of a school principal has been a focal point of discussion and research interest in the recent years. Identifying the core functions of school principals and developing training strategies are considered the corner stones of the reform efforts in response to the rapidly changing social and economic demands (Li, 2004).

<u>Principals as Government Officials</u>

Two types of school principals existed in the long Chinese education history, regardless of historical and political changes: one of an educator and the other of a social activist (Chen, 2004). However, the belief that school principals are representatives of the governing authority is deeply rooted in the Chinese value system. For centuries, the feudal dynasties maintained and perfected a system of keeping educational entities

closely attached to the political regime. The central government's control was especially

stringent since the establishment of new China in 1949. Schools were considered the

"battle ground" of winning the communist ideology war and principals were appointed by

the local government based on their political alignment with the party doctrines (Li,

2004).

Administratively, the principals received government civil service rankings.

Their appointments, responsibilities, promotion and evaluations were based on similar

rules that apply to other government officials. Stipulations established in the 30 years

after 1949 clearly indicated that the school principals' main responsibility is to follow the

instruction of the local government, which received its guidelines from the central

government in Beijing. Loyalty in implementing government and party policies was the

key area of concern according the stipulations. Although clearly outlined, the purpose

and methods of managing schools, ensuring education quality, and developing the

channels of working with teachers, students and parents were outdated and misleading

(Huang, 2004).

The system of treating school principals as typical government officials had 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).

As the Chinese society gradually opens to the global market and with the momentum of economic reform, the tight control over the education system is being relaxed. 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 has pushed the redefinition of the role of school principals to the top of the educational policy agenda (Chu, 2003).

School Principals as Professionals

The first sets of policy papers calling for developing professional school principals occurred in 2000 (Li, 2004). Since then, several research centers with national reputations and leading education policy researchers initiated a number of research projects, some funded by the Chinese Ministry of Education (MOE), on principal professionalization. The leading Chinese educational newspapers sponsored seven rounds of expert panel discussions on the roles and functions of school principals in 2004, taking the research to a new level.

The review of available literature on Principal Professionalization shows that the change of a school principal is twofold: first the evolvement from government official to manager, second the elevation from a generic management occupation to a specialized profession.

Classifying school principal as a management occupation was the first step to rid

school leaders from the political linkage that the position carried in the past. Scholars

that advocate this transformation emphasize that school principals should be treated the

same as other career occupations such as doctors, teachers, or lawyers. These protections

have to satisfy certain occupational requirements including qualifications and

performance standards (Wang, 2003). As a management occupation, principals will be or

should be required to have the knowledge and skills to "manage" – to plan, organize,

communicate, motivate and to evaluate (Huang, 2004). There is also a proposed

threshold of qualifications for the occupation. Desirably, school principals should have at

least college degrees in teaching-related majors, with several years of teaching experience,

and are certified by authorized principal training entities (Huang, 2004). It is also

believed that belonging to an occupation also gives legitimacy and occupational identity

to the principals, which in turn brings self-imposed discipline and accountability (Wang,

2004). With this evolution, principals no longer are considered appointees of the central

government but part of the labor market in which they can be hired, fired and have to

compete based on their qualifications.

However, because schools are unique, with an education mission that impacts the

public good, managing a school is fundamentally different from managing a for-profit

business (Chu, 2006). This is where the occupational definition of the school principal is

differentiated from other occupations.   The principalship, the researchers argue (Huang,

2004), is an occupation that focuses on education as its main responsibility area. The

purpose of education is to develop enlightened individuals. Therefore the effectiveness

of a school is not measured by the amount of profit, but the quality of education in terms of student "body and soul" development (Li, 2004).

Advocating for principal professionalization became a major policy and research force shortly after discussion on the occupational definition of school principals began. Started in 2002, several major papers were published by Hongqi Chu of Beijing Normal University, who lead his research team on a Ministry of Education funded project on Principal Professionalization and Training (Chu, 2003). These papers were considered milestones in the field that elevated the policy discussion from defining the occupation of the principal to specifying professional standards and necessary systemic support for a new generation of school principals.

The key difference between occupation and profession is that professionals are those who have received specialized education and training, equipped with field-specific knowledge and skills, and practice their occupations based on a set of collectively recognized standards (Larson; 1977; Collins; 1979; DiMaggio & Powell; 1983; Elmore, 2007). Eight elements of the principal profession were called for in the Chinese context: (a) long-term specialized training; (b) comprehensive knowledge base; (c) respected ethic code; (d) clear professional practice guidelines; (e) rigorous qualification requirements; (f) professional autonomy; (g) relatively high social esteem and income level; and (h) established professional organizations (Chu, 2003).

Establishing a Knowledge Base for Principal Professionalization: Imported Educational

Leadership Theories in China

To professionalize school principals in China, two main challenges exist: providing training for preparation and development, and implementing support systems for school leaders. Recent scholarly discussions and policy considerations have addressed these two challenges (Chu, 2006). Establishing a knowledge base is essential to provide meaningful and effective training for preparation and development, Due to historical and political reasons, domestic theories and empirical evidence for educational leadership are limited in China. As a first step, researchers and educators turned their attention to "Western" countries for reference.

Here we heed the need for more precise and discriminating use of language when using the term "Western." Some writers use terms such as "Western," "Eastern," or "Asian" in drawing comparisons casually with little attempt to define or distinguish these labels, a serious omission when there is likely to be as much variation within each of them as between them (Dimmock, 2000). For example, major contextual and cultural differences apply between English-speaking Western countries such as the UK and the USA, let alone between the UK, USA, France and Germany, with their different languages and locations on two continents. In this dissertation, the term "Western" refers to Greco-Roman and Judeo-Christian cultures (Wong, 1998) in Europe and North America.

Within Western nations, the conceptualization of principal leadership has evolved considerably since the late 1980s. Predominant notions of the principal's role have evolved from manager, to street-level bureaucrat, to change agent, to instructional

manager, to instructional leader, to transformational leader. Within the past decade, there has been a discernable shift in emphasis in the conceptualization of the principal's role (Hallinger & Leithwood, 1998). Studies from the early to late 1980s were dominated by an instructional leadership conceptualization drawn from the effective-schools literature (e.g., Murphy & Hallinger, 1988). More recently, less emphasis has been given to instructional leadership role and more to models that explicitly acknowledge the implications of school restructuring, such as transformational leadership (Leithwood & Jantzi, 1993). This evolution of the educational leadership role has been labeled as reflecting increased attention to "second-order" changes (Leithwood, 1994) as it is aimed primarily at changing the organization's normative structure.

Western theories and school leadership models have been gradually introduced to China and were frequently referenced since the late 1990's as the education system was going through its transformation, facing the constantly changing societal and economic landscape discussed earlier. Among them, transformational, instructional leadership, distributed leadership, and moral leadership appear to be some of the most prominent strands of theories noticed by the Chinese.

Traditionally, China's education was very highly centralized. Since 1985, China has delegated more authority to local governments and schools. In 1985, the Central Committee of the Communist Party of China (CCCPC) issued the *Decision on the Reform of Chinese Educational System*. More specifically, it was stated that government control of schools was too rigid and management inefficient and that authority should be "devolved" to lower levels. In 1993, the *Guidelines for the reform and development of education in China* was issued by CCCPC and the State Council. This reform document

20

provides enough room for local levels to take more responsibility for basic education both in terms of management and finances.

There has been a major shift of responsibility for decision-making from local officials to schools. Schools have also been encouraged to involve their communities in decision-making processes. Along with the reform, more authority has been given to the principals. However, more authority also means more obligations, more responsibilities and more accountability. The decentralized school system requires principals to have strong leadership and problem-solving skills to guide them in making a broad range of decisions.

Transformational leadership, often labeled as "change-oriented leadership" or "transformational leadership" (Yukl, 2002), was introduced first and also frequently studied in China (Chen, 2006). The spotlight here is on organizational processes (e.g., supporting staff)—employing effective methods for getting the school and its members (staff, students, families, community agents) to become more productive (Leithwood & Jantzi, 2003; Marzano, Waters, & McNulty, 2005). With respect to organizational improvement, transformational leadership focuses on increasing the school's capacity to innovate. Rather than focusing specifically on curriculum and instruction, transformational leadership seeks to build the organization's capacity to select its purposes and to support the survival of changes to the school's core technology. Leithwood and his colleagues at the Ontario Institute for Studies in Education conducted the core theoretical and empirical work on this model. Several of their studies examined transformational and transactional leadership constructs (e.g., Leithwood, 1994; Silins, 1994). The available evidence suggests that meeting the excellence and equity challenge

in urban schools depends on school leaders' action in effectively guiding instructional improvement (Barth, 1986; Leithwood, 1994).

The concept of transformational leadership is attractive to Chinese scholars and practitioners alike because of the volatile nature of the education system as it responds to the external environment and the pressing need for skills and ability to cope with change. Principals now have more responsibilities for school long-term development. They must plan the schools' future ahead. They must learn how to do strategic planning. Some principals cannot adapt to the changes, are not able to manage strategically and independently. They were used to relying on the directives of the government, and to managing day-to-day routines instead of utilizing big-picture type of strategies. This is also the result of the centralized management system and the lack of autonomy at the school level in the past (Chu, 2008). Five challenges in school transformation have been identified. They are: to crystallize values, to distribute authority and build teams, to establish flexible structure, to open communication channels, and to internalize motivations (Bai, 2006).

Following transformational leadership, instructional leadership also received much attention from the Chinese educators. According to Knapp, Copland, and Talbert (2003) "leadership for learning means creating powerful, equitable learning opportunities for students, professionals, and the system, and motivating or compelling participants to take advantage of these opportunities" (p. 12). The core for this strand of leadership include the ability of leaders (a) to stay consistently focused on the right stuff—the core technology of schooling, or learning, teaching, curriculum, and assessment and (b) to make all the other dimensions of schooling (e.g., administration, organization, finance)

work in the service of a more robust core technology and improved student learning (Knapp, Copland, & Talbert; 2003).

Chinese scholars found that the instructional leadership model is particularly meaningful in helping principals focus on student learning, teacher professional development and the self-development of the principal (Bai, 2006; Zhang, 2005). In their view, such focus can be accomplished through several venues: (a) Setting shared vision for the school community, which is different from the traditional top-down approach from the central government. (b) Establishing learning communities among teachers, guidance counselors, and the administrator. This requires leaders in high performing schools to be diligent about providing professional learning information to colleagues on a consistent basis and in a timely manner. In addition, instructionally grounded leaders should monitor the instructional program in its entirety, assuring alignment between learning standards and objectives and classroom instruction. (c) Providing professional development for teachers. In the age of advanced technology and global competition, new instructional tools and methods, information technology and foreign language ability are now emphasized as areas of learning. Effective leaders are especially expert in opening up a wide assortment of improvement opportunities for teachers. (d) Promoting the self-improvement of the principal from past experiences, best practices, and formal leadership training, instead of relying on seniority and previous administrative experiences. (e) Maintaining a flat organizational structure. Compared with the traditional organizational structure, such structure will be essential to establish a learning-centered environment where there is a sense of collegial support, giving teachers the

opportunity to share their views with the principals and to participate in decision-making (Bai, 2006).

Distributed leadership is among the newest concept that is being studied in China (Fang, 2005; Chen, 2006) even though concepts of team leadership and shared leadership have existed in the Western literature for several decades. Leadership literature shows that spreading leadership more generally in an enterprise can help lift the organization to heights that simply cannot be achieved by a single leader (Argenti, 1976; Dubrin, 2004). In education, this more organizationally-grounded and distributed perspective on leadership (Elmore 2000; Spillane, 2006) takes a variety of forms. According to Murphy (2005), on one front, one sees the addition of more formal leadership roles in schools (e.g., teacher coaches and teacher mentors). On a second front, one discerns the spreading of leadership functions and tasks more widely among members of the school community (e.g., a teacher taking responsibility for coordinating the master schedule at a high school). The development of professional communities of practice with significant flows of both formal and informal leadership is the third dimension.

To Chinese school principals, the potential value of distributed leadership is the benefit of sharing power, a relatively new concept distinct from the traditional hierarchical structure (Chen, 2006). With teachers actively participating in curriculum design and instructional improvement, the limited resources that a school receives is extended and enriched. However, because school context varies greatly from school to school, region to region, implementing the distributed leadership model requires locally-driven strategies. The power of the principal indeed has increased along with the educational reform efforts. But often the newly gained school-level authority stops with

the principal. Teachers are not sharing such authority. The democracy concept is still weak among school principals. A 2005 teacher survey found that teachers in schools with lower levels of student achievement tend to perceive their principals with concentrated administrative authority but few responsibilities, as compared with the perception of their teacher counterparts in schools with stronger performance or higher rankings on their principals (Chu, 2005).

Moral leadership, a theory focusing on morality, has taken root in the Chinese education research studies and is considered as having the philosophical underpinning closest to the China tradition and culture (Wong, 1998; Ma & Sun, 2006; Gao, Wang & Lin; 2006). Sergiovanni (1992), among others, argued against those leadership theories that emphasize rationality, logic, objectivity, explicitness, individuality, and detachment. He advocated theory that emphasizes emotions, group membership, sense making, meaning making, morality duty, and obligation. He also critiqued leadership based on technical-rational authority. He argued that teaching and learning are human activities that are too complex to be reduced to rules and procedures, even purportedly scientific ones. Technical-rational leadership assumes that expertise can be reduced to a knowledge base that exists apart from the complexity of the educator as a person and even from the actual practice of teaching. Sergiovanni believed that teaching practice is too idiosyncratic, nonlinear, and loosely connected to student outcomes to be standardized in a manner that lends itself to rational-technical leadership (1992). He contended that a higher level of leadership authority is to be found in the professional and moral domains. When professional authority becomes a driving force, leaders rely on standards of practice and professional norms as reasons to appeal to teachers for action or change

(p.40). He put forward the proposition that people are by nature morally responsive and are capable of responding to duties and obligations that stand above their own self-interest.

The moral leadership proposed by Sergiovanni carries dual implications. The first suggests that the leader must appeal to the followers' sense of righteousness, obligation, and goodness as motivations for action and work. The second implication is that the leader must possess a sense of righteousness, obligation, and goodness himself or herself (Wong, 1998). Without this, it is difficult to conceive how followers could be motivated to follow a moral path. Hence, to argue for a moral dimension of leadership implies a moral standing on the part of the leader.

Chinese culture has a long history of cultivating leadership on moral grounds. When the Chinese turned to the West for modernization in the last century, however, they gradually adopted Western models of social and political organization. Gradually, many traditional values have been rejected or adapted. To understand the intrinsic value system of Chinese leaders and the influence of such value system on their leadership styles, two surveys were conducted on the behaviors and beliefs of Chinese leaders (Lin, 2003; Gao, Wang & Lin; 2006).   The first survey was done in 1988 (n=8,792), and the second survey using the same questionnaire was done in 2003 (n=4,700). The survey used a framework that classifies leadership behaviors into three components: (a) Character and moral; (b) Performance; and (c) Maintenance – thus called the CPM framework (Gao, 2006). An "Implicit Leadership Appraisal Scale," as it was called in Chinese, was constructed that addresses the values and beliefs that people have about the "ideal leader". The scale includes four factors: Goal-achieving effectiveness, multi-capacities, personal

morality and interpersonal competence. Each factor scale includes 19 items (Lin, 2003).

Survey subjects were asked to rank on a scale of 1 to 10 the extent to which they agree

certain leadership quality or ability is necessary for an ideal leader. Researchers found in

1987 that among the four leadership factors, personal morality explains most of the

survey item variances compared with the other three factors. Specifically, items for the

personal morality scale include willingness to serve, honesty, respect for facts,

truthfulness, openness to criticism, altruism, keeping promises, disciplines, no corruption,

and leading by example (Lin, Chen & Wang, 1987). Fifteen years later, with economic

reform, quantum leaps in technology development and seemingly significant social

changes, researchers hypothesized that the value structure would change reflecting the

changes that have occurred in the society. Interestingly, the results show that personal

morality remained as the most influential factor of the intrinsic leadership structure.

Why does personal morality take on such an important role in leadership in China?

It is deeply rooted in the Chinese traditional value system and philosophy traced back to

Confucius and his many followers. It may also be the product of a long history of highly

centralized governance structure where well-established law and management

transparency were lacking. In such situations the morality of the leader in power became

the only hope for equality and fairness (Gao, Wang & Lin, 2006).

In summary, three of the four strands of leadership theories that are widely

discussed and referenced in the Chinese research literature, transitional leadership,

instructional leadership, and distributed leadership, are change-oriented and instruction-

focused, where the action of leaders may be more visible and measurable. Moral

leadership, on the other hand, takes on a different dimension of the leadership domain

that is more intrinsic but appears to play a strong role in how principals are received in their schools.

Academic Learning or Social Learning: Emerging "Learning" Priorities in China and the United States for Basic Education

Since passage of the 2001 Elementary and Secondary Education Act, titled the No Child Left Behind Act (NCLB, 2001), the issue of Adequate Yearly Progress (AYP), the academic proficiency standards that all public schools have to meet annually, has dominated education policy debate. There are those who see adequate yearly progress, or AYP, as the vehicle through which the federal government will finally foster quality education in America's public schools. There are almost as many individuals, at least according to popular polls, who view AYP as the tool opponents of public education will use to dismantle our traditional system of public schools.

Researchers and educators point out that under the NCLB Act, there is a growing focus on student *academic* learning because the core challenge facing America's schools, especially urban schools, is improving student achievement and decreasing the achievement gap. School leadership, especially principal instructional and transformational leadership, is widely recognized as important in promoting processes and conditions such as rigorous academic standards, high-quality instruction, and a culture of collective responsibility for students' academic success (Goldring et al., 2007). The focus on *social* learning, however, is not at the forefront of the new accountability system, at least in terms of performance standards and widely publicized "report cards" (Burroughs et al., 2005; O'Conor et al., 2007; Rothstein, 2004).

Interestingly, a reverse trend is emerging in China. The term that defines this new goal of education in direct translation, is "quality-oriented education." Since the mid 1980's, the central government, through the Ministry of Education (MOE), slowly but steadily incorporated this theme into its educational reform initiatives (State Council, 1993; MOE, 1994; People's Education Daily, 1996). Decades of overwhelming 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 (Ministry of Education, 1999; 2006).

The push for "quality-oriented education" was the guiding principle of the massive curriculum reform of basic education in 2001, which has been carried forward into the current school improvement efforts. Teaching and learning goals that emphasize values and ethics, creativity and independent problem-solving, and citizenship-building were incorporated into the 15-year Strategic Plan for Education of the ninth Chinese National People's Congress (State Council, 2001), which was revisited and confirmed in the Educational Reform Plan of 2005 (MOE, 2005). The new priority has brought on important changes to teacher preparation programs, curriculum structure and content, instructional methods, and most relevant to this dissertation, the ways of measuring school success and school leader effectiveness.

The Learning-Centered Leadership Framework and the Vanderbilt Assessment of

Leadership in Education™ (the VAL-ED)

If leadership is one of the most essential element of school performance, naturally

a key question to ask, is what types of leadership behaviors lead to effective schools? An

assortment of practitioners and academics over the last three decades has helped us see

that not all leadership is equal, that certain types of leadership are especially visible in

high performing schools and school districts (Goldring et al., 2007; Murphy et al., 2006).

Building upon the conviction that instructionally-focused and change-oriented

leadership are especially effective frames for education, supported by the Learning-

Centered Leadership conceptual framework was established to inform the crafting of a

new evaluation system for school leaders and school leadership teams, the Vanderbilt

Assessment of Leadership in Education ™ (the VAL-ED). Focusing on the measuring

leadership job performance—that is, leadership behaviors and practices, the core of the

assessment system is an instrument that measures leadership behaviors. The conception is

aligned with a research-based definition of educational leadership that is rooted in school

improvement (Goldring et al., 2007).

Leadership Assessment: an Essential Element of Principal Professionalization

Principal leadership assessment and evaluation can be an integral part of a

standards-based accountability system and school improvement. 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 accountability for school-wide improvement, leadership assessment can set the organizational goals and objectives for the school leader. When the domains of school leadership that impact student achievement are included as the assessed targets (Goldring et al., 2007; Heck, Larsen & Marcoulides, 1990; Heck & Marcoulides, 1996), leadership assessments help school leadership focus on those behaviors that are associated with student learning.

In the United States, the most important leverage points for improvement in educational leadership are as follows: (1) standards, (2) licensure, (3) program accreditation, (4) professional development, and (5) leader evaluation and resultant consequences (Porter et al., 2006). To date, it is believed that considerable work has unfolded over the first four of these leverage points for redefining school leadership. The widespread development and adoption of national Standards for School Leaders (ISLLC) by 40-plus states and all the major professional associations in school administration, the complete overhaul of program accreditation (the NCATE process) based on those standards, the development of a standards-based national licensure examination by the ISLLC project (the ETS School Leaders Licensure Assessment), and new models for professional development linked to standards are examples of major initiatives.

Unfortunately, only minimal traction has been gained on the fifth critical leverage point, leadership assessment.

In China, the role of principal assessment is gaining more attention in a larger context of pushing for educational reform to improve both the quality and equity of public education. Two important strategies of school leadership development are promoted in China (Chu, 2003). First, "practical knowledge" must be used instead of "theoretical knowledge" in the movement of professionalize Chinese school principals (p.240). Chu recommends establishing a principal leadership framework that is based on what is actually needed in the field for an effective leader. The framework will be constructed by analyzing core job functions and leadership profiles. The existing theories and field-generated leadership descriptions will then be sorted to fit into this framework. It is important, he points out, to distinguish "what should be and what we know is there" in terms of core leadership knowledge and skills (p.241). In other words, the best way for school principals to succeed in the process of professionalization is to link their own experiences to theoretical arguments. The second and more vital strategy is to ensure systemic support for the effort of professionalize school principals in China. Chinese scholars have been refreshingly outspoken about the lack of government structure and resource to support educational reform (Chu, 2006). On the other hand, they are also very adamant about government's role in protecting the public good nature of the education system from privatization and marketization. They argue that because the market mechanism in China is far from mature and the government regulatory entities are still incomplete, exposing the education system completely to market forces and Western education models will be detrimental, especially in ensuring equity in educational

opportunities and establishing stable education leadership (Chu, 2003; Chu, 2006). Five

aspects of the systemic support are discussed: (a) a comprehensive human resource

management structure for principals; (b) a principal licensure system; (c) training for

principal preparation and professional development; (d) principal evaluation; and (e)

incentive and promotional opportunities for principals.

The Learning-Centered Leadership Assessment Conceptual Framework

<u>Definition of Leadership</u>

The Interstate School Leaders Licensure Consortium (ISLLC) Standards for

school leaders (CCSSO, 1996) is widely adopted in the United States as the "golden

standards" of school leadership, in which effective school leaders are described as

"strong educators, anchoring their work on the central issues of learning and teaching and

school improvement. They are moral agents and social advocates for the children and the

communities they serve. Finally, they make strong connections with other people,

valuing and caring for others as individuals and as members of the educational

community" (p. 5).

Built upon the ISLLC standards, the conceptual foundation of the Learning-

Centered Leadership Framework (Murphy et al., 2007) defines leadership as "the process

of influencing others to achieve mutually agreed upon purposes for the organization"

(Patterson, 1993, p.3).   "First, leadership is a process; it is not a personal trait or

characteristic of an individual. Second, leadership involves influence; it requires

interactions and relationships among people. Third, leadership involves purpose; it helps

organizations and the people affiliated with them, in our case schools, move toward

reaching desired goals. This definition of leadership highlights the fact that leadership

can be shared amongst multiple actors and relies on complex, organic interrelationships

between leaders and followers" (Murphy et al., 2007, p. 1-2).

The School Leadership Assessment System Model

The Learning-Centered Leadership Framework establishes a leadership

assessment system model (see Figure 1) that attempts to capture in broad strokes how

education leadership has and might be assessed (Murphy et al, 2007; Porter et al., 2006).

The model has the following important features:
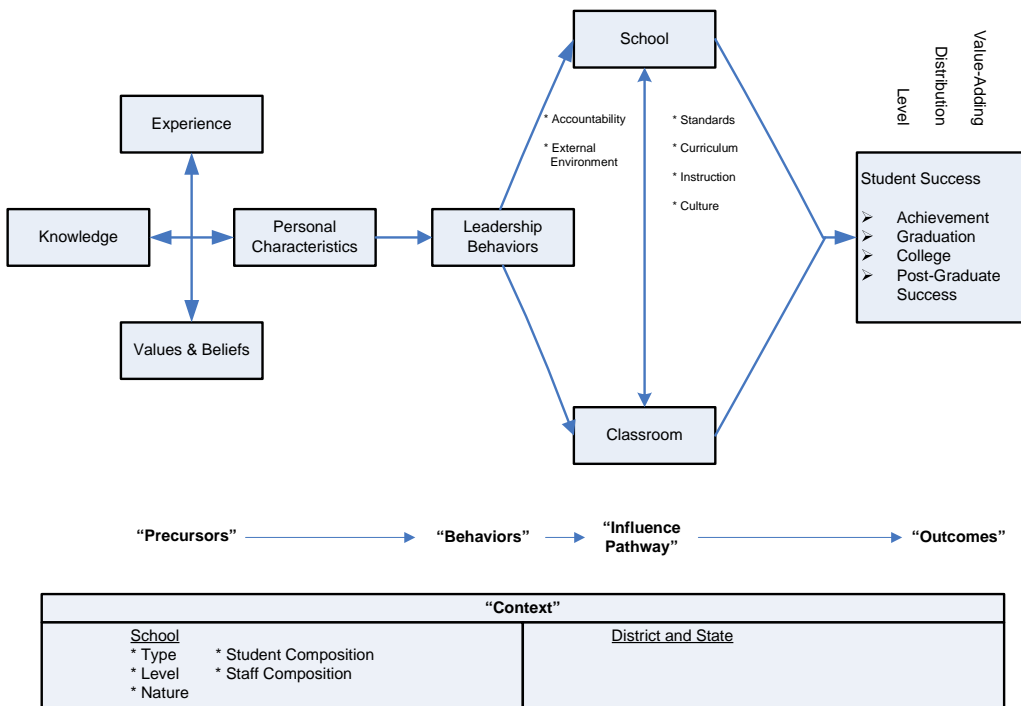
Figure 1: Leading-Centered Leadership Conceptual Framework (Murphy et al., 2007)

1.   The focus of this assessment model is on leadership behaviors. The model

        shows leadership knowledge and skills, personal characteristics, and values and

beliefs as precursors of the actual leadership behaviors exhibited by individuals or teams in performing their leadership responsibilities. The model emphasizes that assessment of education leadership should focus on leadership behaviors found in the literature on effective schools and school districts.

2. <u>Leadership behaviors directly impact the core components of school performance</u> that include standards, curriculum, instruction, culture, accountability, and external environment. The assessment model proposes to give weight to how successful the school is in terms of its core components (e.g., does it have a rigorous curriculum?). These school performances, in turn, lead to student success. Here the thinking is to focus on value-added, for example, improvements in student achievement, student attendance, student graduation rates, and college enrollment. Thus, in assessing a leader or leadership team, one might focus on knowledge and skills, personal characteristics, and beliefs, but that is not the focus. The assessment model does not envision direct effects of leadership behaviors on student success. Rather, the leadership behaviors lead to changes in school performance, which in turn lead to student success.

3. <u>Leadership behaviors impact student success indirectly through the core components of school performance</u>. The assessment model also gives weight to student success measures for the quality of education leadership (e.g., does the school have a relatively large value-added to student achievement?).

4. <u>There are aspects of the context within which leadership and schooling takes place that might moderate the impact of leadership effects</u>. For example,

everything else being equal, the evaluation of leadership quality might

appropriately take into account the experience of the leadership, length of time

in the same school, student body composition, staff composition, level of

schooling, and geographic setting of the school.

<u>Definition of Learning: Student Academic and Social Learning</u>

The focus of leadership efforts, according to Learning-Centered Leadership

Framework, is learning. Here learning is defined by two aspects: academic and social

(Murphy et al., 2007; Porter et al., 2006). Although there are no extensive and specific

discussions regarding "learning" as a construct domain, what "learning" entails can be

inferred from how the outcomes of effective leadership are described. The impact of

leader behaviors in terms of a number of valued outcomes at three periods of time are:

indicators of in-school achievement (e.g., grades on common final exams), measures of

performance at exit from school (e.g., graduation), and more distal indices of

accomplishment (e.g., college graduation) (Murphy et al., 2007). The model also posits

that outcomes be viewed using "a tripartite perspective—high overall levels of student

achievement (quality), growth or gain (value added), and consistency of achievement

across all subpopulations of the student body (equality)" (Murphy, Hallinger, & Peterson,

1986, p. 154).

The weight, as it appears, is relatively heavy on academic learning. Education

quality, evident by student success and school achievement, is largely measured by test

scores, graduation rates, and college entrance ratios; education equality, the other major

mandate on public education, is determined based on the achievement of students from

various backgrounds, in academic subjects. Post-graduation success, assumingly

36

includes social learning, is listed as an element of outcomes, however, it is not fully elaborated. Such strong emphasis on academic learning reflects the national educational policy trend in the United States.

Conception of Leadership Behaviors

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? A comprehensive review of the research literature (see Goldring et al., 2007; Murphy et al., 2007) reveals two key dimensions of highly effective leadership related to student learning and achievement: *core components* and *key processes*. 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 (Conley & Goldman, 1994; Leithwood, 1994). Effective learning-centered leadership is at the intersection of the two dimensions: core components created through key processes (see Figure 2).

Core Components

Core components are linked to student learning and teacher's opportunities to improve their instruction and are aligned with the Interstate School Leaders Licensing Consortium (ISLLC) standards, viewed as a credible and useful foundation for leadership evaluation since they reflect a broad professional consensus on essential leadership domains (Goldring et al., 2007). Not included in the Learning-Centered Leadership Framework are other aspects of leadership such as values and knowledge that while

important, are not behaviorally anchored. The core components of learning–centered

leadership represent the extent to which the principal ensures the school has: high

standards of student learning, rigorous curriculum (content), quality instruction

(pedagogy), a culture of learning and professional behavior, connections to external

communities, and performance accountability (Porter et al., 2006, see Appendix A for

definitions).

Core Components	Key Processes					
	Planning	Implementing	Supporting	Advocating	Communicating	Monitoring
High Standards for Student Performance						
Rigorous Curriculum (content)						
Quality Instruction (pedagogy)						
Culture of Learning & Professional Behavior						
Connections to External Communities						
Systemic Performance Accountability						

Figure 2. Learning-Centered Leadership Framework Conception Chart (Porter et al., 2006)

Key leadership processes refer to the ways in which leadership, individually and collectively, influences organizations and their constituencies to move toward achieving the core components (Murphy et al., 2007; Porter et al., 2006). Such processes, according to the Framework, include: planning, implementing, supporting, advocating, communicating, and monitoring (See Appendix A for definitions).

The leadership behaviors to be assessed are defined at the intersection of these two dimensions as in Figure 2 (Porter et al., 2006). According to the conception, for example, one would "assess the extent to which the school leadership plans for a rigorous curriculum (the intersection between planning and rigorous curriculum) or implements high-quality instruction (the intersection between implementing and high quality instruction)" (Porter et al., p.4).

The Vanderbilt Assessment of Leadership in Education ™ (the VAL-ED): *The Instrument*

The VAL-ED is a paper and on-line assessment that utilizes a multi-rater, evidence-based approach to measure the effectiveness of leadership behaviors known to influence teacher performance and student learning. The VAL-ED is also a "360 degree" assessment: teachers, the principal, and the principal's supervisor respond to the behavior inventory. Both *core components* and *key processes* in the leadership conception are measured at the intersection of the two dimensions. The outcomes of the assessment include a behavior inventory or profile, interpretable from both norm-referenced and standards-referenced perspectives (Porter et al., 2006).

The VAL-ED requires respondents to make judgments about a principal's leadership behaviors that influence teachers' performance and students' learning. Respondents are specifically asked how effective the principal is at specific actions that effect core components of learning-focused leadership. The effectiveness ratings range from 1 = Ineffective to 5 = Outstandingly Effective for each of 72 behaviors (see Appendix B-a and Appendix B-b). These behaviors sample all 36 cells of our conceptual model of leadership equally and thus serve as indicators of the construct of leadership the Framework desires to measure (Elliott, 2008).

The respondents are asked to rate the extent to which the principal ensures behaviors and actions are taken in the school, thus acknowledging that principals do not necessarily perform the behavior themselves, but often designate and distribute these leadership practices and behaviors throughout the school. Respondents' ratings of effectiveness should be based on evidence they have collected or reviewed during the current school year. If a respondent does not have any evidence upon which to make an effectiveness rating, he/she must rate the principal as *Ineffective.*

When completed as intended, a leader can earn a total raw score ranging from a low of 0 to a high of 360 on each of the three respondent forms of the VAL-ED. These raw scores will be transformed to normalized standard scores (e.g., mean 100, standard deviation 15) and percentile ranks. The standard scores will be presented within a 95% confidence interval. The higher the score, the more effective a principal is perceived as exhibiting the desired attribute of leadership. The end goal of the interpretation is to be able to make a reliable and valid attribution about a principal based on the input from

multiple respondents who have observed and interacted with him/her over the course of a school year.

Establishing Psychometric Properties

Development Steps and Validity Questions for a Technically Sound Assessment

The design of the VAL-ED is directly influenced by technical standards for high-quality assessments (*Standards for Educational and Psychological Testing,* AERA, APA, & NCME, 1999), principles of universal design (NCEO, September 2006), and time-tested practices of item and test development (Downing, 2006; Haladyna, Downing, & Rodriguez, 2002). Collectively, these professional documents and the published research on test development and high-quality assessment of human performance provide strong guidelines for designing a high-quality and successful assessment program for school leaders.

The development of the VAL-ED has been guided by a comprehensive plan that involves: (a) specifying the purposes of the assessment, (b) defining content assessed, (c) writing items, (d) developing test specifications for validity evidence plans, (e) designing instructions and response format, (f) piloting test forms, (g) designing scoring and interpretation frameworks for scores, (h) conducting studies that yield evidence for the reliability and validity of the scores, (i) refining items, format, and score interpretation procedures, (j) field-testing forms with a representative sample, (k) developing norms and standards to guide interpretation of results, and (l) writing a technical manual that summarizes technical characteristics and sound uses of the assessment (Elliott et al., 2008).

Item and response scale development : A total of 144 (72 each for two parallel survey forms) were selected from a pool of more than 200 items written based on review of learning-centered leadership literature and alignment to ISLLC standards. The items were critiqued by education leaders and leadership researchers.

Item sorting study: The sorting study served as a first step in testing the validity of the assessment measure. The study sought to establish a level of content validity, identifying whether the items within the instrument measure the domains that they were constructed to measure. Nine principals were recruited to the task. Each was provided with the definitions of each core component and each key process and the 36-cell matrix in Figure 2. A pool of 294 items was divided into three random sets stratified by cell. Overall, the results of the sorting study indicate that, at least for school principals, the behaviors captured by the 294 items appear content valid when judged against the conceptual framework of core components by key processes against which the items were written.

Cognitive lab interviews: Cognitive labs were designed to augment psychometric measures of validity and reliability with more qualitative measures to be sure that respondents are interpreting questions in the same way, or that the full range of appropriate responses is captured (Ouimet, Bunnage, Carini, Kuh, & Kennedy, 2004). The cognitive interview required respondents to "think aloud" as they work through a questionnaire, providing the researcher with a play-by-play of their cognitive processes. Two rounds of cognitive interviews have been completed with the VAL-ED. The importance of the cognitive effort to provide evidence was weighed against the fatigue it

caused and the forms were shortened to 72 items. Feedback was received about the response scales and specific item phrasing. Overall, the response was that the instrument was inclusive and it captured key leadership behaviors.

Initial factor analysis and internal consistency estimates for pilot study sample: In March 2007, a nine-school pilot of the VAL-ED was conducted in a large urban district in Kentucky (Porter et al., 2008). Three each of elementary, middle, and high schools participated in the study, and schools were randomly assigned to forms A and C. Six of nine schools had response rates greater than 75%, and the overall response rate, 72.5%, was roughly the same across forms. Analyses of the pilot data included descriptive statistics, correlations, factor analysis, and estimates of internal reliability. Results suggest that the instrument is feasible. Few item responses were missing (1.9%) or marked "don't know" (7.4%) across teacher respondents and even less so for principals and supervisors, suggesting that respondents understood the items. Sources of evidence were indicated by most all respondents. Effectiveness ratings were positively skewed. School-level means scores indicate that principals gave slightly lower ratings than the teachers; supervisors gave slightly higher scores. Results also suggest high reliability and validity. Cronbach's alpha for each of the item scales was greater than .90 and was slightly higher for core component scales than key process scales. Correlation analyses revealed a positive association between the score a principal gave him/herself and the scores the teachers gave him/her ($r = .47$). Confirmatory factor analyses produced consistently high fit statistics (GFI >.95) for a multifactor solution that directly paralleled our theoretical framework of core components and key processes. The response stem and the scale were revised to address the skewness and to enhance the clarity of responses.

The Imperative Need for Valid and Reliable Leadership Assessment

for the Identification and Development of School Leaders

Assessing the effectiveness of school principals has been an important element of

school improvement for more than two decades. Ideally, a principal assessment should

be easy to administer, capture the essence of the role of a school principal, and provide

valid and reliable data for purposes such as professional development and performance

evaluation. Criticism exists, however, regarding the adequacy of assessment instruments

and the processes employed to evaluate school principals (Heck & Marcoulides, 1996;

Porter et al., 2006; Portin et al., 2006; Reeves, 2005). In fact, as early as 1990, in a

comprehensive review of the literature related to principal evaluation, Ginsberg and

Berry (1990) found a wide array of practices reported with little systematic research to

support one approach over another. In 1992 and 1993, the weakness of research on

school leadership evaluation was the topic of two full issues of *the Peabody Journal of

Education,* in which Ginsburg and Thompson (1992) lamented "the state of research on

principal evaluation emphasizes the lack of empirically supported information about best

practices" (p.67).

The State of Leadership Assessment in the United States

The stakes for effective school leaders are high in today's climate of system-wide

accountability where American public schools are charged with the tasks of improving

student achievement and closing performance gaps among the subgroups of an

increasingly diverse student population (Catano & Stronge, 2006; Portin et al., 2006;

Thomas et al., 2000). Although the rhetoric of making changes to schools is hardly new,

never before have the effectiveness of schools been so closely monitored and measured by quantifiable standards across schools, districts and states. Despite increasing attention on improving the leadership of school principals, and renewed emphases on training and preparation programs, leadership assessment and evaluation has received far less attention and research.

Existing "Western" research on school principal leadership assessment has focused on two areas: leadership dimensions, (i.e., what to assess); and assessment methods and their validity, (i.e., how to assess).

<u>Dimensions of Principal Leadership</u>

To do their jobs well, principals carry out multiple responsibilities, both internal and external to the school environment. Here is how a principal at a Chicago public school describes her daily work: "After a day in which I was part cafeteria manager, registrar, disciplinarian, social worker, procurement officer, nurse, human resources officer, and chief financial officer of a multi-million-dollar budget, I took some time to reflect on the primary job I have ahead of me this year: being the instructional leader of a school that must raise its test scores by 10 percentage points across the board, or face increased sanctions under the federal No Child Left Behind law" (NPR, 2007).

Because of the complexity of the principal's role, the main difficulty in the field of school principal leadership assessment is identifying the leadership dimensions that should be assessed. (Glasman & Heck, 1992; Hart, 1992; Huff, 2006; Marcoulide, Larsen & Heck, 1995; Oyinlade, 2006). Four approaches to what to assess have been suggested: (a) responsibilities, (b) knowledge and skills, (c) processes, and (d) organizational outcomes.

Approach one is based on specific job tasks or lists of responsibilities (Ginsberg, 1992). Job tasks associated with the principalship generally include the responsibilities for managing school programs, pupil personnel, community relations, physical facilities, student behavior, and coordinating professional development. This approach has been widely used throughout the 20th century prior to the presence of high-stakes testing and systemic accountability.

Instead of focusing on key roles or tasks, another approach has been to use key competencies, knowledge and skills that principals should possess (Thomas, Holdaway & Ward, 2000). Olyinlade (2006) presented a method of assessing school leadership effectiveness using 18 items of "essential behavioral leadership qualities" (p.32). Examples of the 18 items include good listening skills, good presentation skills, and participative decision-making style. Although these items typically measure knowledge, skill and abilities instead of "what the principal does," the author argues that the content of the instrument is based on what the field and experts perceive as essential "behavioral qualities" for an effective principal.

To respond to the concerns of not fully covering effective leadership domains, some researchers promote the approach of using "effective school correlates" or best practices that emerged from research on the effects of principals' activities on school improvement as the framework for deciding what principal behaviors to assess (Hallinger & Murphy, 1987; Heck, Larsen & Marcoulides, 1990). This approach focuses on the "process" through which leadership affects school-wide performance. For example, *the Interstate School Leaders Licensure Consortium* (ISLLC) includes a set of "components of professional practices" for each of the six leadership standards (CCSSO, 1996).

Assessing the "process" side of school leadership can be very difficult and complex. Some researchers and school systems focuses on organizational outcomes. Outcome-based evaluation focuses primarily on desired school outcomes and the degree to which the school has been able to achieve these outcomes, (e.g. increased student achievement, higher attendance, lower drop outs). Although this approach seems to be better aligned with performance accountability and has received a significant amount of attention, it faces methodological hurdles (Heck & Marcoulides, 1996; Rowan, Raudenbush & Kan, 1991) especially in assuming direct causal relations between what the principal does and school outcomes. Further, relying solely on outcome based assessment runs the risk of ignoring organizational and contextual factors that can help explain student achievement and other outcomes.

Assessment Procedures

Beyond the difficulties related to "what to assess" is the challenge of determining appropriate methods to establish the assessment process and to make valid inferences on principal performance. Several survey studies have provided snapshots of principal evaluation procedures and extended their reach much closer into the actual practices of principal leadership assessment and evaluation in schools and districts (see Lashway, 2003). For example, from a survey of 800 principals in Ontario, Canada, Leithwood and Montgomery (1986) identified problems in appraisal practices such as the lack of detailed policies for the process, the standards of performance were not always well publicized, and the practices outlined in policies were not followed.

Lashway (2003) noted several studies of evaluation practices, including a study by National Association of Elementary School Principals (Doud & Keller, 1998), which

47

found that the evaluations were most often carried out by central office personnel, although respondents reported a growing trend to involve parents, teachers, and principals themselves. A study in 17 California districts (Lashway, 2003; Stine, 2001) identified three types of evaluations in use: Checklists rating principals on a variety of behaviors or traits, ranging from time management to loyalty; free-form evaluations consisting of a narrative and measures of principal performance against a set of predetermined goals. These different formats were often combined resulting in a wide variety of procedures. A nationwide survey by Reeves (2005) found that principals agreed that their evaluations were generally positive, accurate, and consistent with job expectations. However, fewer found the evaluation process relevant to enhancing their motivation and improving their performance. The respondents of the survey also indicated that their evaluations lacked the specificity to indicate what behaviors should be changed. Based on the survey results, Reeves (2005) also noted that most principals reported not having received useful feedback from their evaluations, assessments were inconsequential, and the criteria of evaluation were unclear.

Taken together the state of the knowledge base regarding quality, use, and influence of principal leadership assessment is limited.   In a review of leadership assessment in education, Portin et al. (2006) point out that the broad trend of increasing emphasis on learning and school improvement in the recent decades has made an impact on what and how leaders are assessed. Five shifts, according to the review, merge as the new directions of leadership assessment, namely the movement toward assessing behaviors instead of traits, relying on professional standards, focusing on learning results, emphasizing leadership development, and considering organizational context. The

evidence of such shifts, however, is yet to be substantiated by further empirical work on

"the evolving nature and uses of leadership assessment approaches" (p.26). Our review

of the literature, however, found no comprehensive survey of current principal leadership

assessment practices in the field for the recent decades. Without analyses of the content,

format, psychometric properties and usage of the actual instruments, assumptions about

how principal leadership assessment can serve as an important part of the school

improvement equation remain untested.

Wide Range of Variation and Lack of Focus on Learning-Centered Leadership Behaviors

A comprehensive review of current principal leadership assessment practices in

the United States (Goldring, E., Cravens, X. C., Murphy, J., Elliott, S. N., Carson, B., &

Porter, A. C., 2008) analyzed both the general content and the usage of 65 actual

instruments used by districts and states and provided an in-depth look of what and how

districts evaluate their school principals. Using the Learning-Centered Leadership

Framework (Porter et al., 2006), the review focused on identifying the congruency (or

lack thereof) between current evaluation practices and the research-based criteria for

effective leadership that are associated with school performance.

For content, the iterative and deductive method of analysis showed that districts

focus on a variety of performance areas when evaluating their principals. The awareness

of school principal leadership assessment and evaluation varies greatly among the

districts and states. Some districts and states have comprehensive evaluation systems

readily accessible online. In other district and states, school personnel were unable to say

how school principals are evaluated. Approximately one-third of the sampled districts

contacted could not identify an appropriate person for us to speak with about principal

evaluations. The number of items in the collected assessment instruments range from fewer than 10 to more than 180. A majority of the instruments (75%) have fewer than 50 items. Items in short instruments have the tendency to be generic, using categorical terms such as instructional management, school morale, personnel management or administration and fiscal management. Items that are included in the longer instruments are usually specific.

The instruments also vary as to their content emphases. From the analysis of specific subcategories, a wide spread of assessed areas of very limited depth.was found among the sampled principal assessment instruments. Despite the understood view that instructional leadership and leaders should be the focus of the principalship, there is great variation of what is assessed.

When comparing the content of the principal assessment instruments to the core components of the Learning-Centered Leadership Framework, it was found that the critical behaviors that principals perform to influence student achievement are not receiving enough emphasis. For example, the extent to which the principals ensure that the school has a *rigorous curriculum* and *quality instruction*, two very important areas for learning-centered instructional leadership, receive a relatively small share of the items on existing leadership assessment instruments. The core component of ensuring that the school has a *culture of learning and professional behavior*, receives the most emphasis in current assessment instruments. In contrast, only 5% of the items in the average assessment instrument measures the principal's behavior related to ensuring the school has a rigorous curriculum and similarly few items (only 7% on average) focus on principals' engagement with the quality of instruction in the school.

According to the Learning-Centered Framework, *culture of learning and professional behavior* indicates that "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" (Porter et al., 2006, p.4). It is not surprising to see this component emerges as the primary focus of principal assessment practice. It has long been realized that the educational environment of American public schools is most strongly influenced and brokered by teachers (Coburn, 2004; Schwille, 1982; Weick, 1976). Schools with effective principals tend to have higher levels of professional community which in turn leads to higher student achievement. Studies show that school leaders help develop professional community through their attention to individual teacher development, and by creating and sustaining networks of conversation in their schools around issues of teaching and learning (Bryk, Camburn, & Louis, 1999; Bryk & Driscoll, 1988; Louis, Marks, & Kruse, 1996). The sampled instruments reflect the attention given to this component in the field practice, as least in leadership assessment, for principals to play a central role in the extent to which a school exhibits a culture of learning and integrated professional communities.

Brought on by federal legislative mandates such as No Child Left Behind and the ever looming global competition, it is critical in the United States that leaders establish *high academic standards* and *systemic performance accountability*. Increasingly, principals are being asked to ensure that there are individual, team, and school goals for rigorous student academic and social learning by aligning school activities with local, state and federal standards. Furthermore, leaders must hold themselves and others responsible for realizing high standards of performance for student academic and social

learning. In other words, there must be individual and collective responsibility among the professional staff and students and this accountability should be evident in principal assessment instruments.

The analyses indicate that current principal evaluations are not focusing on some of the most powerful indicators for improving student learning: ensuring *rigorous curriculum* and *quality instructions*. What students are taught is a powerful predictor of student achievement on a test (Gamoran, Porter, Smithson, & White, 1997), and it helps explain a portion of the achievement gap between White, Black, and Hispanic students (Porter, 2003). Ideally, teachers teach what is described in content standards. As much as curriculum and instruction are considered as classroom teachers' territory, it is the responsibility of the school principal to ensure that there is ambitious academic content provided to all students in core academic subjects and there are effective instructional practices that maximize student academic and social learning.

In seeking information on how principals are evaluated, it was found that in most cases, the practices of leadership assessment do not align with the *Personnel Evaluation Standards* in terms of assessment utility and accuracy. Most concerning is the lack of clear documentation that aligns with these important personnel evaluation quality measures. Little discussion of psychometric properties, evaluation procedures and evaluator training can be found among the sampled assessment instruments. Information provided by the 44 instrument s indicates that assessments for principals are conducted very differently by school districts, with no clear norms or performance standards. There is little consistency in how the assessments are developed, which leadership standards are used, and if the measures are reliable and valid.

The State of Leadership Assessment in China

As an integral part of the systemic support, leadership assessment is gaining increasing attention in China as a significant "condition of leadership" impacting leaders' behavior. An effective principal evaluation system is a key component of professionalization. However, there has been little evidence of effective practices of using formative or summative assessments to measure and develop leadership knowledge and skills (Zhao & Wang, 2007).

Analysis of School Principal Job Functions

To achieve the transformation of school principals, researchers must find out the current job functions of school principals and how much change has already occurred in the field during the years of economic development and social change (Qiao, 2003).

Findings from several empirical research studies conducted in the last five years indicate that the main functions of the school principals focus on three aspects: planning, fundraising, and attending social activities (Li, 2003; Li, 2004; Qiao, 2003). Instructional leadership, notably, was not among the key functions. Researchers also found that the principals do not report having some of the authority that the Educational Law of People's Republic of China had stipulated such as hiring and firing new teachers (Qiao, 2003). As to qualification requirements and preferences, experiences are more valued than degrees and certifications.

Based on interviews and open-ended questionnaires (Qiao, 2003), researchers at Beijing Normal University (BNU) sampled 150 principals, 100 of which attended the principal training at BNU, and 50 principals from Guizhou Province, a rural region. Despite the descriptive nature of the survey, a substantial amount of qualitative data was

collected on knowledge and skills reported as necessary and important by the principals. The profiles focused on several aspects of the principal's role. The first aspect is the core functions of school principals. Principals report that the majority of their time is spent on dealing with issues outside of the classrooms, most of which are related to securing funding for the schools. This is different from the experience of public school principals in the United States where school funding is based on district formula for per student expenditure. The second aspect, principal autonomy, compares the principal authority that is listed in the educational law and what principals in the field actually have. For example, the law gives the principals the authority to hire qualified new teachers and also the right to have the final say on which students meet the graduation requirements. But according to the survey and interviews, such decisions are usually not made at the school level and frequently interfered with by outside entities. Third aspect is on the knowledge and abilities necessary for the job. Principals ranked the importance of knowledge base in the following order: management, psychology, education theories, law, computer skills, philosophy and one specific academic subject. For necessary abilities, the order is planning, coordination, decision making, research, public speaking, interpersonal communication, writing ability, instructional ability and public relations. The fourth aspect, qualification requirements, shows that a majority of the principals value the experience from teaching more than the experience of being administrative officials. Typically, as reported by the survey and interviews, principals that rise from the director of instructions positions tend to be more effective than those that rose from assistant principal positions, which in China are of administrative nature. The fifth aspect on the typical principal's career path, however, indicates that assistant principals are more likely

to be promoted to be principals (Qiao, 2003). This may be an indication that administrative experiences carry more weight than instructional experiences in the evaluation process.

<u>Measuring School and Principal Effectiveness</u>

There has not been any specific governmental stipulation regarding principal evaluation except for a recommendation made by the Ministry of Education in 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 used at provincial, city, and township levels as the guidelines for principal evaluations. However, the extent to which such dimensions are covered and the formats used vary greatly (Zhao & Wang, 2007).

Despite the significant growth of research in the area of principal development, studies focusing on principal evaluation have been few (Zhao & Wang, 2007). Among the limited number of published research articles on principalship in mainland China since 1994, 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.

As a part of the Principal Professionalization and Systemic Support project funded by the Ministry of Education, researchers of Beijing Normal University embarked on a three-year study of evaluating the current state of principal assessment and making recommendations for a new system that is meaningful and practical (Chu, 2003). A review of existing national and local policies and practices highlights inadequacies in

both the standards for principal effectiveness and how such standards are operationalized by the districts and schools as assessment tools (Zhao & Wang, 2007). The review points out that the current standards make no distinction among behaviors, results or abilities of the principals when setting assessment criteria, thus causing inevitable confusion and duplication. The four aspects of school leadership recommended by the 1992 MOE document, values, abilities, diligence, and achievement, is a good example of such jumble of different orientations of measuring principals' work. In surveying the actual principal assessment practices, researchers found three main areas of concern. First, using teachers' feedback to evaluate the school principal relies on the important assumption that the collective assessment by the teachers objectively and proportionally reflects the effectiveness of school leadership. However, such assumption is often doubted by school principals and their supervisors for reasons such as the disinterest of teachers in the process or distrust of confidentiality. Second, the current assessment indicators do not reflect the changes that have occurred in schools and districts in the recent decades. Educational reform initiatives that promote character-building, all-around skill-enhancing and curriculum upgrade are not included in the existing measures. Third, other important constituents of the school community especially those that are traditionally considered external such as parents and entities that make investment in the school system are not included in the assessment process. Only having internal members, principals, supervisors, teachers and staff, for the evaluation of principal leadership yields an incomplete picture, some argue (Zhao & Wang, 2007).

In summary, research and practices in principal leadership assessment suffer from two deficiencies: First, the lack of a sound theoretical framework for principal leadership

that links the objectives of education with leadership standards. Second, the lack of an

assessment system that is developed with empirical research evidence, and is valid and

reliable (Chu, 2003; Zhao & Wang, 2007).

Challenges of Cross-Cultural Comparison and Adaptation

of Leadership Theories and Their Applications

Care must be taken in validating conceptual constructs such as leadership across

cultures. Not only may the particular leadership framework being emphasized vary

culturally, but the same framework may have different meaning within different cultures

(Heck & Marcoulides, 1996). 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, 1997; Hallinger & Kantamara, 2000; Heck & Marcoulides, 1996).

The American Educational Research Association (AERA), American

Psychological Association (APA), and National Council on Measurement in Education

(NCME) *Standards for Educational and Psychological Testing* (1985) provides careful

directions for educational measurement specialists and psychologists who select, develop,

administer, and use educational and psychological tests. Three of the standards in this

publication are especially relevant in the context of test adaptation:

> *Standard 6.2.* When a test user makes a substantial change in test format,
> mode of administration, instructions, language, or content, the user should
> revalidate the use of the test for the changed conditions or have a rationale
> supporting the claim that additional validation is not necessary or possible.
> *Standard 13.4.* When a test is translated from one language or dialect to
> another, its reliability and validity for the uses intended in the linguistic
> groups to be tested should be established.

Standard 13.6. When it is intended that the two versions of dual-language tests be comparable, evidence of test comparability should be reported.

These standards provide a framework for considering sources of error or invalidity that might arise in efforts to test the fit of an assessment from one language and culture to that of another culture with different language and leadership traditions. High level of test adaptation sophistication is seen today in both TIMSS and Organization for Economic Cooperation and Development's Program for International Student Assessment (OECD/PISA, see, e.g., Grisay, 2003; Hambleton, 2002). Sources of error or invalidity can be organized into three broad categories (Hambleton, Merenda, & Spielberger, 2005): (a) cultural/language differences, (b) technical issues, designs, and methods, and (c) interpretation of results. Each source of error will be discussed in light of its implication on this dissertation and how concerns may be addressed by proposed studies in the following chapter for research methods.

It is important to point out that although much of the research literature on theoretical and methodological issues and guidelines are more specifically related to academic and psychological tests such as student achievement or personality inventory, the discussion is very relevant to other forms of assessment seeking cross-cultural comparison as well.

Cultural and Language Differences

It is important that the examination and interpretation of cross-cultural results consider all parts of the assessment when considering the cultural and language differences (van de Vijver & Leung, 2000; Hambleton, Merenda, & Spielberger, 2005) beyond the narrow context of just the translation or adaptation of tests. For example,

Western theories may not fit in the "cultural dimensions" of the Chinese society (Gao, Wang & Lin, 2006), where the tradition of centralized governance, emphasis on personal morality, and male-dominance tendency may hamper the implementation of Western models that rely on decentralized approaches and clarity in laws and regulations. More importantly, it is very difficult to implement the leadership models without systemic change and system-wide support, which call for reform at the policy level. For example, instructional leadership with innovative, student-centered pedagogy has little chance to succeed under the immense pressure of the current entrance examination structure. Students, teachers, and parents are forced to focus on preparing for the exams for survival instead of true quality of education.

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, Merenda, & Spielberger, 2005).

Construct Equivalence

Determining if construct equivalence exits between different cultures under study is a prerequisite for doing any cross-national, cross-cultural, or cross-language comparisons. Construct equivalence encompasses both conceptual and functional equivalence as well as equivalence in the way the construct measured by the test is operationalized in each language/cultural group (Harkness, 1998).

Determining whether construct equivalence exists between two cultures involves judgmental strategies. To be able to ensure conceptual/functional equivalence and equivalence of construct operationalization, several approaches were suggested by Van

de Vijver and Poortinga (2005) and Sireci, Patsula, and Hambleton (2005), including interviewing or observing people from the cultures of interest, researching the cultures of interest, and asking others who know about the cultures. These ways are subjective, and therefore, the use of multiple sources of evidence is highly recommended.

Test Administration

Communication problems between a test administrator and examinees can pose a serious threat to the validity of test results. Perhaps the test directions are not clearly communicated because of adaptation problems. One way to circumvent problems, but not always feasible, is to ensure that the instructions on the test itself are clear and self-explanatory, with minimal reliance on verbal communication (van de Vijver & Poortinga, 2005). 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 will 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 the iterative improvement process described previously.

The proper selection of test administrators can be helpful too. Hambleton et al. (2005) recommend that they should (a) be drawn from the target communities, (b) be familiar with the culture, language, and dialects, (c) have adequate test administration skills and experience, and (d) know the importance of following any standardized procedures associated with the test. For example, 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. Additionally, consistency in test administration across different groups can be improved by providing (basic) training to all test administrators.

Item Formats Used

There are also concerns over the differential familiarity with particular item formats that may 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 United States for achievement testing, it cannot be assumed that everyone is as familiar with multiple-choice items (Hambleton et al., 2005).

The Influence of Speed on Examinee Performance

The original concern brought up by Hambleton et al. (2005) is that item and test bias that may occur due to the role of test speededness because not all cultural groups have the same experiences with speeded tests. Although not quite applicable with the case of principal assessment, how much time teachers, principals and their supervisors have to fill out the survey and the possible bias caused by rushing should be evaluated.

Technical Issues, Designs, and Methods

Five technical factors that may influence the validity of tests adapted for use in other languages and cultures (Hambleton et al., 2005). These are: the test itself, selection and training of translators, the process of translation, judgmental designs for adapting tests, and data collection designs and data analysis for establishing equivalence. Among these factors, some factors are more relevant than the others to this dissertation

The Test Itself

It is ideal if the test is developed with its possible use in a different cultural setting in mind, which may reduce problems later in the adaptation process (Hambleton & Patsula, 1999). However, it is not the case with the VAL-ED in the Chinese context.

Therefore, care was taken to choose vocabulary and expressions that both maintained the original construct meaning and were easily understood across the languages and cultures.

Selection and Training of Translators

Experts emphasized that translators should be more than persons familiar and competent with the languages involved in the translation (Hambleton, Merenda, & Spielberger, 2005). They should know the cultures very well, especially the target culture (i.e., the culture associated with the language of the adapted test). This knowledge is often essential for an effective adaptation. Also, subject matter knowledge in the adaptation of achievement tests is highly desirable. The nuances and subtleties of a subject area can be lost on a translator unfamiliar with the subject matter. Too often, translators without technical knowledge resort to literal translations that are often problematic to target-language examinees and threaten test validity. Finally, test translators would benefit from some training in test construction. A test translator without knowledge of the principles of test and scale construction could easily make test material more or less difficult unknowingly, and correspondingly, lower the validity of the test in the target population.

The Process of Translation

Concerns over choosing the correct dialect within a language, using similar frequency counts of words, and finding ways to deal with nonexistent equivalent word for translation were discussed regarding this factor. Because both English and Chinese are very well developed languages with large vocabularies and similar grammatical structures, these concerns were addressed.

Judgmental Designs for Adapting Tests

The two most popular designs, forward translation and backward translation, were discussed by Hambleton et al. (2005). Between the two, the back-translation design is the best known and most popular of the judgment designs (Widenfelt, 2005). In its most popular version, one or more translators adapts a test from the source language (English in the case of the VAL-ED) to the target language (Chinese for this study). Different translators take the adapted test (in the target language) and adapt it back to the source language. Then, the original and the back-translated versions of the test were compared and judgments are made about their equivalence. To the extent that the two versions of the test in the source language look similar, support is provided for the equivalence of the source and target versions of the test. The back-translation design can be used to provide a general check both on the quality of the translation and to detect at least some of the problems associated with poor translations or adaptations. Researchers especially like this design because it provides them with an opportunity to judge the original and back-translated versions of the test so that they can form their own opinions about the adaptation process. This is not a possibility for them with the forward-translations design unless they are proficient in both languages.

Data Collection Designs and Data Analysis for Establishing Equivalence

For large-scale test comparison efforts, there are three data collection designs commonly used to evaluate the equivalence of the factor structure of the test and of the items (or rating scales) in difference languages. These designs are: (a) Bilingual examinees take source and target versions of the test; (b) source-language monolinguals take the original and back-translated versions; and (c) source-language monolinguals take

source language and target-language monolinguals take target language. These designs, however, were not be practical considering the scope and the relatively limited resource for this dissertation.

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 should 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 U.S. and Chinese settings.

In short, based on the review of literature on issues, designs and guidelines of cross-cultural adaptation and comparison of tests, it was imperative for this dissertation to first ensure that construct equivalence was established between the original VAL-ED instrument and the translated version. The dissertation was aimed at 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. It is also very important to carefully

choose test administrators and use appropriate item formats. With regard to interpretation of scores, specific background variables that impact on performance need to be carefully considered. In this regard, differing leadership standards, levels of motivation, and socio-political factors were especially important. It should be kept in mind that comparisons were not be undertaken only with emphasis on the differences. Similarities between language and cultural groups also provided useful and relevant information.

CHAPTER III

RESEARCH METHODS

A series of studies were designed, address methodological concerns that are specifically related to the cross-cultural translation of the VAL-ED and to collect substantive evidence of construct validity and score reliability for the VAL-Ed. Three studies, described in the sequence in which they were conducted, addressed the major research questions. Each study embodied sub-studies that contributed to one or more research questions (see Table 2).

Specifically, Research Question 1, *the fit of the theoretical framework*, was answered mainly with the expert-panel examination results from Study 1, but Study 2 and Study 3 also produce qualitative and quantitative evidences from examining the instrument itself that gauged the fit of the theoretical framework. Research Question 2, *the validity and reliability of the VAL-ED scores,* was addressed mainly by Study 2 and Study 3, using both qualitative and quantitative data to evaluate content, test-criterion validity, and instrument reliability. In addition, information collected from Study 1 regarding the theoretical framework served as an important aspect of the validity check for the VAL-ED scores because the instrument was designed to conceptualize the Framework. Research Question 3, *the relevance and utility of the framework and the instrument,* relies on the composite results from all three studies. For this question, whether and the extent to which the frame and the instrument converged with priorities and standards of Chinese education were discussed, and preliminary recommendations

were made on whether the existing framework and the instrument should be modified, and if yes, what and how.

Table 2: Studies and Sources of Validity Evidence

Research Question 3	Research Question 2	Research Question 1	Study	Sub-Study	Sources of Validity Evidence	Proposed Sampling and Data Collection
Relevance and Utility of the Framework and the Instrument (Study 1, 2, 3)		*Fit of the Theoretical Framework (Study 1)*	1	Alignment Analysis	Evidence based on assessment content (Content Validity)	Expert-panel: principal training faculty (5 from Beijing Normal Univ., 3 from South China Normal Univ.; 2 Bureau of Ed. Officials; 2 principals)
		Validity and Reliability of the VAL-ED Instrument (Study 2, 3)	2	Cognitive lab interviews	Evidence based on assessment content (Content Validity)	2 supervisors 2 principals (elementary, middle, high) 2 teachers (elementary, middle, high)
			3	(1) Factor Analysis	Evidence based on internal structure (Construct Validity)	19 schools: 18 principals, 1165 teachers 6 supervisors of the 18 principals
				(2) Internal Consistency Check	Evidence based on internal structure (Reliability)	
				(3) Performance Nomination Summary	Evidence based on relations to other measures (Criterion Validity)	

Study 1: Expert-Panel Examination of the Learning-Centered Leadership Framework

and the VAL-ED Rating Scale

The objective of this first study was to examine the *content validity* of the

Learning-Centered Leadership framework and the VAL-ED instrument items, addressing

mainly Research Question 1: How well does the Learning-Centered Leadership

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? The study had two parts: alignment analysis and translation

modification for the framework and the instrument.

Participants of the expert-panel (N = 12) were comprised of five faculty members

of Beijing Normal University, three faculty members of South China Normal University;

two officials of provincial bureau of education; and two principals in principal

professional development training. They were selected as individuals that were

knowledgeable in the subject area and could provide valuable feedback regarding current

professional standards and practices of school principals in China. The selection was

made based on the recommendation of two visiting scholars to Peabody College that are

leading researchers in the field of school principal development and educational policy,

from Beijing Normal University and South China Normal University.

Alignment Analysis: The alignment study was designed for four content

validation purposes: first, to find out how much of the content of the VAL-ED items

collectively and individually reflect the leadership practice and standards in Chinese

urban schools; second, to see if there were any missing core components or key processes

in the Chinese context, e.g. moral leadership; third, to find out if one or more core

components or key processes might be considered not "core" or "key" in the Chinese

context; and lastly, to detect any differences in the interpretation of the core components

and key processes for further clarification and modification in the translated version.

The expert panel members were first given the concept paper (Porter et al., 2006)

on the framework and the VAL-ED assessment, both in Chinese. Each member was

contacted via email and encouraged to read the documents, especially the conception of

the important elements (core components and key processes) and their definitions.

Each member was given an *Alignment Rating Form* (see Appendix C) and asked

to examine the LCL Framework by: (a) rating the alignment of the VAL-ED items with

current principal practices (relevance) and the priorities and standards for Chinese urban

schools and principals (importance) on a Likert scale from 1-5; (b) indicating if and what

core components or key processes that were missing; and to write down whatever might

be considered missing in the blank space provided.

The rating forms were distributed and collected via email. For *importance*,

participants were asked to rate the importance of the leadership behavior to student

learning and school success, on a Likert scale of 1-5, where 1=not important at all, 2= not

very important, 3=somewhat importance, 4=quite important, and 5=very important. For

*relevanc*e, participates were asked to rate the extent to which the same leadership

behaviors were currently incorporated into the principals' work, with 1=little to none, 2=

a little, 3=somewhat, 4=much, and 5=very much.

Mean ratings on *importance* and *relevance* were calculated for each item and

analyzed at various levels: by core component, key process, and rater type. Criteria that

gauge the ratings on importance and relevance were decided based how the rating scale

was constructed. For example, to count how many items of leadership behavior were considered important to student learning and school success by the participants, a rating of 3.0 served as the cutoff for ratings beyond neutral. Similarly, to calculate how many leadership behavior items were at a below-average level for their relevance in the principals' actual workload, a rating of 3.0 was used as the cutoff point. Comments and suggestions were also analyzed and summarized.

Translation Modification: Based on results from the alignment analysis, translation of the instrument items was refined for better clarity and improved reflection of the construct. To ensure that the existing LCL Framework and VAL-ED items could be tested for cross-cultural fit, no components, processes, and items written for their intersections were deleted or added.

Study 2: Validation of Instrument Content Construction

The purpose of this second validity study was to assess the construct validity of the instrument before the assessment items were finalized by evaluating the construction of the instrument content, including both the assessment items and the design of the instrument upon completion of the expert-panel examination of the framework.

Cognitive lab interviews were the method used to conduct this study. The principals were identified from participants of principal training programs offered at South China Normal University in fall of 2007. The list is available to the public upon request. Invitation letters were sent to all (n=50) principal participants. In the letter, the purpose of the research was fully explained, along with issues of risks, benefits and confidentiality. Five principals volunteered. The voluntary principals were also asked to

70

nominate 1 or 2 teachers from his/her school to participate in the studies, and to let the researcher know if his/her supervisor might be contacted to participate in the cognitive interview. The teachers and supervisors were contacted separately by the researcher via letters and follow-up phone calls. Upon working out scheduling conflicts, a total of six participants (two principals, two teachers, and two supervisors of principals) eventually participated in the interviews.

Cognitive labs augmented psychometric measures of validity and reliability with more qualitative measures to ensure that respondents are interpreting questions in the same way, or that the full range of appropriate responses is 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.

The cognitive interview methodology was used to identify respondents' problems with the instrument's instructions and to check on the type of evidence used during the evaluation process. Feedback on translation, interpretation, evidence use, scale use, and time taken to do the assessment were recorded and analyzed. The interviews also focused on several areas that might be of concerns according to the pilot studies of the VAL-ED in the U.S. (Porter et al., 2008): (a) if the amount of effort many respondents put into the sources of evidence raise the cognitive demand of the task; (b) if the respondents periodically forget the stem, "the principal ensures the school…", focusing instead on whether the principal himself or herself actually performed the behavior; and (c) if the respondents defer to outcomes regardless of the key process the item was seeking to highlight, while they need to rate each item based on "behaviors".

The cognitive interviews were designed to provide opportunities for addressing inadequate translation due to contextual, linguistic and other reasons when the participants talk through their thinking process. All aspects of the assessment, the instruction, the definitions in the framework, the items themselves, the listed sources of evidence, and the rating scale, were included in the cognitive interview protocols (Appendix D). Again, the purpose of changes to the wording and the sentence structure were not to alter the original construct intent of the items but to ensure construct equivalence, an essential condition of cross-cultural construct validation.

The process of examining the content validity of the items was also designed to yield important insight regarding the extent to which construct equivalence is achieved after the LCL Framework elements (core components and key processes) and the items were translated into Chinese. Clarifications, questions, and suggestions made by the expert panel on the Chinese version of VAL-ED items were taken into consideration. Modifications were made so that vocabulary choices and sentence structures of the translated items could better reflect the original intent. As discussed previously regarding the challenges of cross-cultural comparison, achieving construct equivalence between the original and the translated assessment instrument must go through an iterative process during the study sequence, in which the expert panel, the cognitive interviews each contributed to the refinement of the final version of the translated VAL-ED instrument.

Study 3: Studies of the Reliability and Validity of VAL-ED Scores

The final version of the translated VAL-ED instrument was used to examine the validity and the reliability of the framework and the instrument using rating scale scores collected from a sample of principals, their supervisors, and teachers in their schools. Analyses of the VAL-ED ratings included descriptive statistics, correlations, factor analysis, and estimates of internal reliability.

To avoid possible coercion from the local bureau of education (comparable to the 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 offered at South China Normal University in the fall of 2007. Letters were sent to all principal participants. The purpose of the research project was fully explained, 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 with the variation in academic ranking, economic condition of the school zone, school size, and school type (elementary, middle, high) in mind. Such school profile information is publicly available through the Guangzhou City Bureau of Education, where South China Normal University is located. 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 1165

teachers from 19 schools, 18 principals from the same 19 schools, and six supervisors for the 18 principals.

Table 3 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 includes a good range of schools in Guangzhou in terms of school type, percentage of senior teachers, geographic location and school size. There were seven elementary school (37%), three high schools (16%), three middle schools (16%), four schools that are have both middle and high school grades (21%), and two vocational schools (grade 10-12). Teacher classification is based on years of experience and instructional quality, and the 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% to as high as 90%. The schools also varied in size, from small elementary schools with fewer than 500 students to large high schools of more than 2000 students. One vocational school had more than 5000 students. The schools were located in six different school districts among the 10 that were in the city of Guangzhou. Some of the districts were in the downtown metropolis Guangzhou, and some were in the outskirt of the city considered very rural and agricultural less than 15 years ago. Many of the student families are migrant workers who work 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 1 year to more than 10 years. The principals had an average of 3.7 years of tenure at the existing schools.

The overall return rate for the teacher assessment is 97.4%, and school-level return rates are above 90% for all the schools in the sample. The return rate for the principals is 95%, with one principal out of 19 did not return the assessment. The return rate for the supervisors is 100%: all assessments evaluating the 18 participating principals were returned by the supervisors.

Table 3: Profiles of Sampled Schools

学校 ID	School Type	# of Teacher Assessments Issued	# of Teacher Assessments Returned	Teacher Assessment Return Rate	Total # of Students	Pct of Senior Teachers	Years as Principal	Years As Principal of This School
1	Elementary	55	55	100.00%	1629	60%	12	6
2	Vocational	87	83	95.40%	5300	20%	8	8
3	Elementary	68	67	98.53%	780	53%	2	2
4	Elementary	65	59	90.77%	1468	71%	2	2
5	Vocational	27	27	100.00%	2000	15%	1	1
6	High	153	149	97.39%	2686	32%	5	2
7	Middle & High	121	120	99.17%	1355	38%	3	2
8	Middle & High	35	35	100.00%	783	13%	6	2
9	Elementary	36	35	97.22%	669	48%	5	5
10	Middle & High	36	35	97.22%	780	13%	6	5
11	High	60	56	93.33%	2783	46%	7	7
12	Elementary	41	40	97.56%	1012	66%	2	2
13	Middle	47	47	100.00%	1000	10%	3	2
14	Middle	51	45	88.24%	1174	10%	9	9
15	Middle & High	155	155	100.00%	4500	21%	6	4
16	High	33	33	100.00%	2200	32%	7	6
17	Elementary	30	30	100.00%	436	45%	10	2
18	Elementary	44	43	97.73%	1193	90%	2	2
19	Middle	52	51	98.08%	2742	27%	NA	NA

After obtaining informed consent, the translated VAL-Ed (including the new items from the previous studies) were administered to supervisors, principals and teachers,

using paper-and-pencil format. Participants were asked to review the assessment instructions (provided on the first page of the assessment) and complete the assessment (one form, based on U.S. VAL-ED Form A). All participants were given adequate time (one hour or more) to complete the assessment. No personal identification was recorded on the assessment form to link the result to an individual.

There were three analytical components of this study, each used the same set of data collected from assessing the same group of participants in the same process, which include effectiveness ratings of the 72 VAL-ED items, and the ratings on a set of four items for Chinese leadership standards. The analyses included a factor analysis, two types of reliability analyses, and a performance nomination analysis.

The dataset was examined to determine the amount and patterns of missing data. When item data were missing and the pattern appears random, a conservative approach was used, estimating the missing values by inserting means for the items. When more than 10% of the items in any case are missing, the case was not used in the analyses because the mean estimate procedure reduces the variance of the overall score for these cases. Percent of missing data was reported in the chapter for results.

Factor Analysis. An exploratory factory analysis (EFA) was conducted to examine the factorial validity of the VAL-ED items in the Chinese setting. EFA is usually used to explore the number of factors that account for the co-variation between variables when there is no a priori sufficient evidence to form a hypothesis about the number of factors underlying the data (Stevens, 1996).

For the U.S. pilot study, a confirmatory factor analysis (CFA) was used to evaluate the structure of the VAL-ED and its fit to the theorized multi-dimensional model

of learning center leadership. The original VAL-Ed items were assigned to subscales (core components and key processes) specified based on the justification of the theoretical framework. The factor analytic model for this justification is illustrated in Figure 3. In this four-order factor model each item is expressed as a function of the intersection of one core component and one key process. Each of the 36 latent variables representing a cross of six core components and six key processes are to be reflected by two items. The 36 second-order factors are expressed as a function of six core components or six key processes. These six third-order factors are expressed as a function of a single fourth-order factor representing the overall leadership effectiveness score.
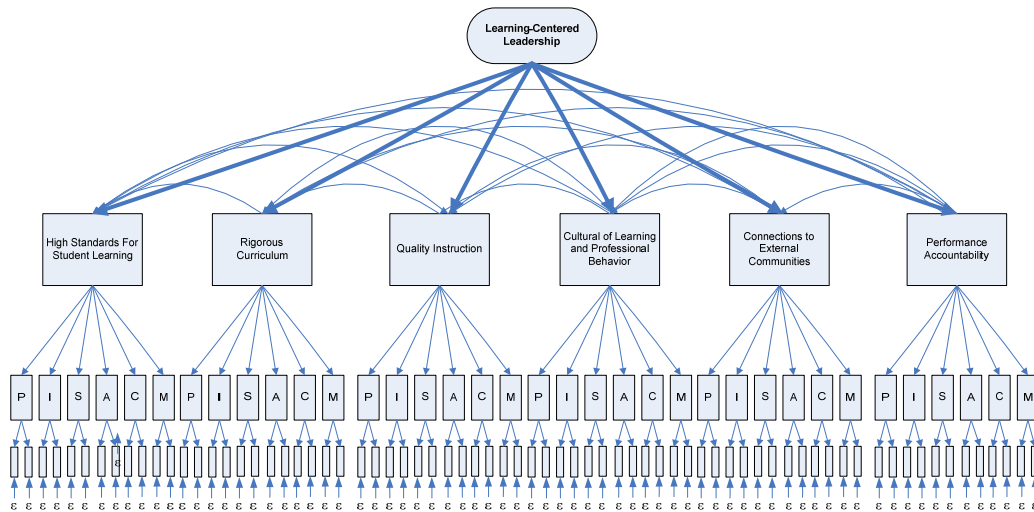
Figure 3. Factor Structure Illustration of the VAL-ED Items

It was determined pre-mature to make the same theoretical assumptions about the VAL-ED's structure when the framework and the assessment instrument were applied in a different cultural setting. Therefore, it was important to conduct repeated procedures of factor analysis without setting the number of factor criteria until evidence of content

validity and discriminatory potential can be identified to fit the empirical results <u>and</u> the conceptual framework (Schechter, 2008; Thompson, 2004). The results from the expert-panel analysis of the framework and the cognitive interviews provided preliminary qualitative information on the fit of the framework and the instrument content, and further confirmed the need for EFA to explore the factor structure of the translated VAL-ED.

With this original factor structure in mind, the degree to which the core components and key processes posited in the VAL-ED measurement framework as reliable latent constructs in the observed data was evaluated using exploratory factor analysis. Ideally, this is an analytic process of exploring the factor structure that needs to be repeated until a preliminary instrument with both conceptual meaning and reasonable measurement characteristics is achieved. Such factor structure then can be compared with the structure that was established in the U.S. setting.

The data for this analysis were responses from the effectiveness rating items collected from all of the assessments. The sample size, 1165 teachers in 19 schools, was sufficient for a minimum factor loading of .40 (Hair, Anderson, Tatham, &Black, 1998). An EFA of the item matrix was performed to study which VAL-ED items clustered together and which did not. 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 focuses 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 varimax criterion (Kaiser, 1958) was used to examine the 72-item instrument.

Reliability Estimate Studies. Reliability is an essential part of any test or assessment and concerns the *consistency* of the scores. Two methods are used to estimate the reliability of VAL-ED scores. First, Cronbach's alphas 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. Cronbach's alpha (Cronbach, 1951) will generally increase when the correlations between the items increase. For this reason the coefficient is also called the internal consistency or the internal consistency reliability of the test.

The inter-rater reliability measures check on 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 .29 were considered low; correlations $\geq$ .30 to .59 were considered moderate; and correlations > .60 were considered high.

Performance Nomination Summary. This part of the study was intended to measure criterion validity of the LCL Framework and VAL-ED, comparing 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 are currently used as broad categories of principal

evaluation criteria in China (MOE, 1992; Zhao & Wang, 2007). The rating scale is also

on a Likert scale of 1-5 and defined by the same rating levels used in the VAL-ED for

equivalence comparison. The four dimensions are listed as four items, separately from

the VAL-ED items (See Appendix B-c). The sample for this study was the same as for

the factor analytic study.
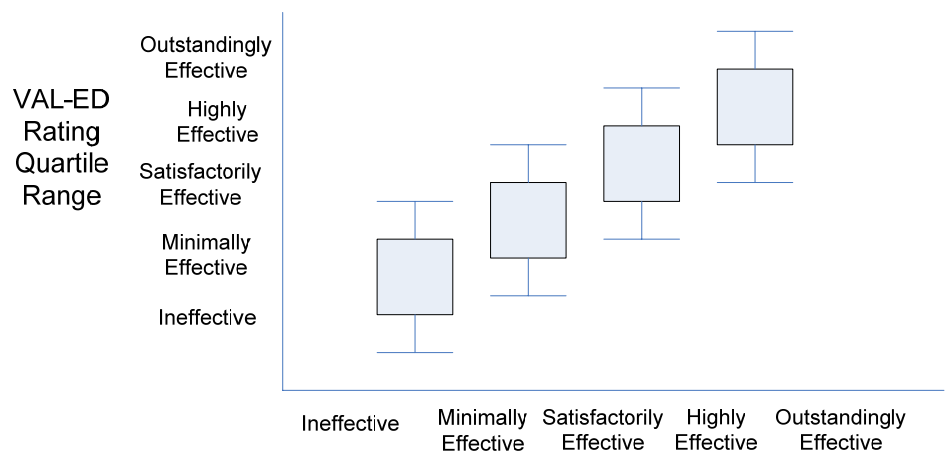
Performance Range Based on
the Four-Dimension Chinese Leadership Standards (MOE, 1992)

(Hypothetical Depiction of Perfect Convergence)

Figure 4. Performance Nomination Summary

The dispersion of mean scores was analyzed to measure the extent to which the

instrument reflects the current principal evaluation criteria in China. The box-and

whisker plots (as illustrated in Figure 4) summarizes a hypothetical spread of ratings on

the Chinese four-dimension standards, as reported by the respondents on the same principals that also are rated by the VAL-ED scales. If the two sets of criteria converge perfectly, the quartile ranges of the VAL-ED effectiveness total scores (average of the 72 items), depicted by box-and-whisker plots, should reflect the continuum of the performance categories for the personal perception of the four-dimension standards, demonstrating the extent to which the construct domain covered by the Learning-Centered Leadership Framework coincides with how effective leadership is perceived in the Chinese setting.

Addressing Cross-Cultural Issues in the Studies

Significant amount of attention was devoted to the technical issues related to instrument translation. Care was taken to choose vocabulary and expressions that both maintained the original construct meaning and could be easily understood across the cultural and language settings. For the materials used in the studies for this dissertation, a three-step procedure was used for the translation of VAL-ED to ensure linguistic equivalence and etic-free conversion: initial translation, back-translation by a second party, and independent evaluation by a third party familiar with both languages. The three-step translation included all documents that needed to be presented to the participants in the three studies: Information regarding the Learning-Centered Leadership Framework, the complete assessment instrument VAL-ED, recruitment letters, and consent forms. The most important and difficult task was to accurately translate the definitions of the core components and key processes of the Framework, and the 72 items of the VAL-ED instrument. To test cross-cultural validity, translated versions need to

maintain their construct equivalence in the process. In this case, elements of the framework and items in the VAL-ED, once translated, needed to be interpreted by the users in China to reflect the original intent of the English version.

The construct equivalence of the assessment items was evaluated and improved continuously through out the first and second studies until the instrument was finalized and administered. In the initial stage of the translation for documents to be used in the first study, specific obstacles were noted and were taken to the first two studies for further deliberation.

The first examination of the construct equivalence occurred when an expert panel was formed to examine the Learning-Centered Leadership Framework and its two-dimension conception. The panel brought the desired multifaceted perspective on the construct domain. Members of the panel include faculty members that were directly involved in designing and providing principal training from two major teaching universities, bureau of education officials that had jurisdiction of overseeing local schools and school principals. Questions raised by the panel members regarding the definition and interpretation of the framework elements were used to improve the instrument, which was further evaluated during the cognitive interview process. The respondents contemplated upon which category an item belonged, and discussed their thought process toward each item. They provided insight into how the translated items were interpreted and if the understanding is the result of equivalent construct domain. Upon analyzing the cognitive interview results, more adjustments to the translations were made to improve the construct equivalence of between the translated version and the original instrument.

The assessment administration was lead by the author of the dissertation, who meets the four selection criteria set forth by Hambleton et al. (2005) for a suitable test administrator as one that is familiar with the culture, language, and has adequate test administration skills and experience.

The instrument format and possible influence of time in the survey administration were also taken into consideration in the studies. Because the VAL-ED assessment uses a five-point Likert effectiveness rating scale that was familiar to Chinese educators (Zhao & Wang, 2007), the instrument format was proven to be a non-issue. Participants of the studies were informed how they will be engaged in the research activities in the recruitment letter and prior to the activities begin. Sufficient time for each activity was secured through planning with the assisting local university.

Summarizing the design of the studies, this paper used samples of urban educators—teachers, principals, and supervisors—to provide tests of the claim that (a) the Learning-Center Leadership framework and its instrument the VAL-ED measures learning-focused leadership behaviors, and (b) its results are deemed useful by principals and others interested in leadership development in Chinese urban schools. In building the argument for the validity of scores from the VAL-ED, the studies provided a comprehensive view of quantitative and qualitative evidence that encompassed the construct being measured and the utility of the results of the measurement.

CHAPTER IV

RESULTS

The results for both the scale development studies and the initial usage of field

test studies are covered in this chapter. The findings of each study are presented

separately, including tables, figures, data analyses, and noteworthy observations made

during the sampling process and the administration of the studies.

Study 1: Expert-Panel Examination

The expert-panel examination of the Learning-Centered Leadership Framework

and the VAL-ED rating scale, was conducted via email communication with the 12

selected panel members. Among the 12 participants, four were practitioners that work in

the local school systems and eight were researchers in the field of school leadership

development for basic education (first grade to 12th grade). Among the four practitioners,

two are directors of district-level bureaus of education, and have direct responsibilities

for managing educational services of local schools; two are principals, one from an

elementary school, and the other from a high school. Among the eight researchers in

leadership development, two head up principal training centers at their universities, four

are faculty members that teach at university-run principal training centers; two are Ph.D.

students in educational administration with experiences in conducting large-scale

assessments on school principal evaluation and leadership development.

Relevance, Importance, and the "Reality Gap"

The intent of this study was to find out how the content elements of the Learning-Center Leadership Framework and the assessment are aligned with the practice and standards in urban Chinese schools. The alignment analysis took two angles. The relevance dimension gauged the extent to which the leadership behaviors of the VAL-ED were 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. In other words, the relevance ratings are for what it *is in reality*; the importance ratings are for what the standards *should be*. The difference between the rating given to importance and the rating to relevance, is the perceived "reality gap". This alignment approach provides three sets of comparison measures on the fit of the VAL-ED framework and items: how the VAL-ED is aligned with the current practices, how the VAL-ED is aligned with necessary standards for learning-centered leadership, and how the current practices are different from such necessary standards.

Relevance and importance ratings were done by the panel members for each of the 72 VAL-ED items that describe learning-centered leadership behaviors. The ratings were analyzed first using the mean scores of core components and key processes, then at the individual item level.

Relevance and Importance of the VAL-ED Core Components and Key Processes

Table 4 shows the mean scores of relevance and importance on the VAL-ED leadership behaviors. We see that all of the relevance ratings for the six core components and the six key processes are below 3.5. Among them, Connection with External Communities and Systemic Performance Accountability received ratings were lower than 3.0, a level that is a below average; Quality Instructions received a 3.46 average, which is the highest. On the other hand, the core components all received ratings close or above 4.0 on average on importance, indicating that the leadership behaviors are important to the success of schools and students in the eye of the expert panel but may not have been carried out to the same extent in practice. When the 72 items are sorted by the key processes, the same can be said about the difference between the mean scores for relevance 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 relevance vs. importance show a correlation of .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 relevance ratings (3.08) at a highly statistically significant level (p <.001, df=5). For the key processes, the correlation is .90 between the mean ratings for importance and relevance and the two sets of ratings are different at a statistically significant level (p <.001, df=5).

The fact that the mean ratings for relevance are consistently lower than the importance ratings for all six core components and all six key processes of the VAL-ED

indicates the presence of gaps between what were considered necessary to enhance

learning-centered school results and what were believed to be in practice by the

principals.

Table 4: Average Ratings for Relevance and Importance of the VAL-ED Items

Component	High Standards	Rigorous Curriculum	Quality Instruction	Culture of Learning	External Communities	Performance Account.
Relevance	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*	1.09	0.9	0.94	0.88	1.32	1.12
Process	Planning	Implement.	Supporting	Advocating	Comm.	Monitoring
Relevance	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*	0.96	0.91	0.83	1.31	1.17	1.08

*Paired t-test for the difference between the group means is significant at $p<.001$, df=5.

The sizes of the "reality gap" vary among the six core components and the six key

processes (Table 4). Connection with External Communities has the largest difference

between the mean scores of importance and relevance (1.32), meaning while it is

perceived as important (at 3.74), it carries a small weight of the workload (2.41) of the

principals. Rigorous Curriculum and Cultural of Learning show the smallest gaps

between perceived importance and weight of current workload (0.90 and 0.88

respectively). Among the key processes, advocating has the largest "reality gap" (1.31),

while supporting has the smallest (0.83).

The views on relevance and importance of the VAL-ED leadership behavior items

are very similar between the practitioners and the researchers on the panel (see Table 5),

with t-test for paired two-group means showing no statistically significant differences. This could be an indication that the researchers are familiar with the field work of the principals, and their view of the importance of the leadership behaviors to the success of schools and students are also closely in synch with the practitioners.

Table 5: Relevance and Importance as Perceived by Practitioners and Researchers

Components	High Standards	Rigorous Curriculum	Quality Instruction	Culture of Learning	External Community	Performance Accountability
Relevance						
Practitioners	3.19	2.92	3.44	3.27	2.42	3.15
Researchers	3.31	3.18	3.47	3.38	2.42	2.90
Importance						
Practitioners	4.50	3.85	4.60	4.35	3.65	4.23
Researchers	4.29	3.99	4.29	4.15	3.78	4.01
Processes	Planning	Implement.	Supporting	Advocating	Comm.	Monitoring
Relevance						
Practitioners	3.23	3.09	3.39	2.81	2.96	2.85
Researchers	3.28	3.30	3.48	2.58	3.05	2.96
Importance						
Practitioners	4.29	4.23	4.41	3.96	4.33	4.07
Researchers	4.18	4.09	4.22	3.98	4.12	3.96

Despite the similarity, a closer look at the "reality gap," the extent to which practice lags behind perceived importance, reveals some interesting differences between the views of the practitioners and the researchers (illustrated in Figure 5). The "reality gaps" are larger for the practitioners than for the researchers in four of the six core components: High Standards for Student Learning, Rigorous Curriculum, Quality Instructions, and Cultural of Learning and Professional Behavior. However, for Connection with External Communities, the perception is reversed. Based on these differences, it appears the researchers believed that there was larger reality gap between

the importance of this core component and what is in practice. For the component of Systemic Performance Accountability, the perceived gaps are the same.

The "Reality Gap" (Importance-Relevance)
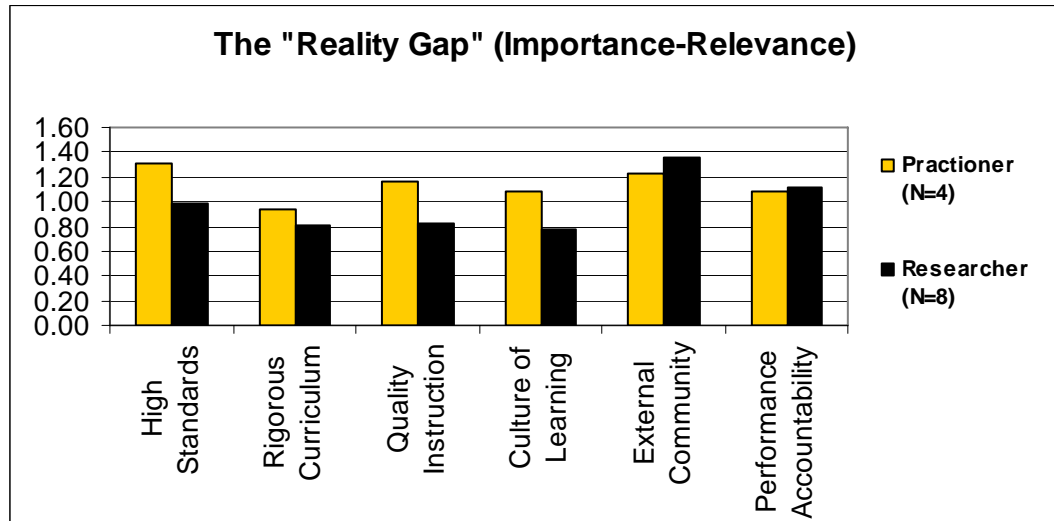
Figure 5: The "Reality Gap" as Viewed by the Practitioners and Researchers

The variation of the "reality gap" among the components and processes is a strong indication that the fit of the VAL-ED may be better in some areas than the others. Better alignment appears to be in the core components of curriculum and instruction, but not in connections with external communities or in the process of advocating. Such view is supported by both the practitioners and the researchers on the expert panel. However, such expectations needed to be substantiated with further inquiries into the nature of the differences.

Importance and Relevance at the Item Level

Analysis of the *relevance* and *importance* ratings was also done at the item level in addition to the mean scores of the core components and key processes. The 10 items

that show the largest "reality gap" between current practice and perceived importance of the leadership behaviors emerge from all core components of the VAL-ED except for Quality Instructions, and all key processes except for Supporting (see Table 6). Even though the "gaps" are in various core components and key processes, the specific content of the 10 items show that the inadequacies appear 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.

Table 6: VAL-ED Items that Have the Largest Gap Between Importance and Relevance

Core Components	Key Processes	Rel.	Imp.	Avg Diff	The VALED Items How effective is the Principal at ensuring the school . . .
	Advocating	2.58	4.50	1.92	#8: challenges low expectations for students with special needs.
High Standards for Student Learning	Communicating	2.42	4.17	1.75	#10: communicates with families and the community about goals for rigorous student learning.
	Planning	2.42	4.25	1.83	#14: plans access to rigorous curricula for students with special needs.
Rigorous Curriculum	Advocating	1.58	3.25	1.67	#19: advocates a rigorous curriculum that honors the diversity of students and their families.
Cultural of Learning and Prof. Behaviors	Advocating	2.25	3.83	1.58	#43: advocates a culture of learning that respects diversity of students.
	Monitoring	2.75	4.42	1.67	#48: assesses the culture of the school from students' perspectives.
	Implementing	2.08	3.75	1.67	#52: builds business partnerships to support social and academic learning.
Connection with External Communities	Communicating	2.33	4.17	1.83	#58: listens to the diverse opinions and needs of all families.
	Monitoring	1.50	3.25	1.75	#60: monitors the effectiveness of community-school connections.
Systemic Performance Accountability	Implementing	2.50	4.08	1.58	#63: uses faculty input to create methods to hold faculty accountable.

These areas of perceived gaps resonate with the difficulties encountered in the translation process. For the initial translation of the VAL-ED items, because students with disabilities, especially those with learning disabilities are taught in separate classes or schools, "special needs" was translated to include students that have difficulties in learning, a broader interpretation. Diversity is another term that when translated, it carries different connotation in the Chinese educational context. While in the U.S. it describes students coming from various socio-economic backgrounds, but more frequently used for differences in race and ethnicity. In China because of the homogeneity of ethnicity (in most urban cities students are 99% Han), diversity usually only covers the socio-economic status of the student families, including the classification of their resident status: urban or rural. Community is another familiar concept in the U.S. educational context. The Learning-Centered Leadership Framework defines external community as "family and/or other people and institutions in the community" (Porter et al., 2006, p. 6). While the translation is clear linguistically, working with the community is a practice not widely adopted in the current Chinese educational setting.

Such gaps between what is perceived as important and what is being practiced were also noted by the panel members, written in the comment space reserved for each item. The two principals commented on Item 19, Item 20 and Item 43, which address the needs of students either lagging behind academically or from diverse backgrounds: "The standardized curriculum makes it difficult for the principal and teachers to attend to the individual needs of students." "It is only possible to help students lagging 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."

Table 7: VAL-ED Items that Have the Least Gap Between Importance and Relevance

Core Components	Key Processes	Rel.	Imp.	Avg Diff	The VALED Items How effective is the Principal at ensuring the school . . .
		3.83	4.08	0.25	#3: creates buy-in among faculty for actions required to promote high standards of learning.
High Standards for Student Learning	Implementing	3.83	4.00	0.17	#4: creates expectations that faculty maintain high standards for student learning.
	Implementing				
	Planning	3.67	4.00	0.33	#13: develops a rigorous curriculum for all students.
	Implementing	3.58	3.75	0.17	#16: implements a rigorous curriculum in all classes.
	Supporting	3.58	4.00	0.42	#17: secures the teaching materials necessary for a rigorous curriculum.
Rigorous Curriculum	Supporting	3.33	3.58	0.25	#18: supports teachers to teach a curriculum consistent with state and national content standards.
Quality Instructions	Planning	4.00	4.25	0.25	#26: plans a schedule that enables quality instruction.
Cultural of Learning and	Planning	4.50	4.08	-0.42	#37: plans programs and policies that promote discipline and order.
Prof. Behaviors	Implementing	3.83	4.25	0.42	#40: builds a culture that honors academic achievement.
Systemic Performance Accountability	Advocating	2.83	3.25	0.42	#67: challenges faculty who attribute student failure to others.

On the other hand, some leadership behaviors were perceived as being practiced at levels comparable to their importance. Among the 10 items that have the least differences between the relevance and importance ratings (Table 7), the focus on

Rigorous Curriculum is most apparent, covering key processes of planning, implementing, and supporting. With the exception of Connections with External Communities, other core components also have items that have the near-equivalent ratings between relevance and importance. Item No. 37 stands out as the only leadership behavior that has a negative "gap" between relevance and importance (Average Importance =4.08, Average Relevance=4.50), indicating a feeling that too much discipline and order were in practice, more than what were perceived as necessary. Such negative gap was reported by both the practitioners (-0.25) and the researchers (-0.50).

The comments of both the practitioners and researchers on the panel offered some explanation on why the differences between the ratings of relevance and importance on leadership behavior items related to Rigorous Curriculum are small. Curriculum content standards are set at the national level in China. A majority of the text books are uniformly produced and distributed with accompanying teaching materials and guides at the provincial level that are aligned with such standards. For this reason, "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, however, rated relevance for Item 23 (evaluates the extent to which all students complete a rigorous curricular program) with a 5, but importance with a 4. He 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 does not apply for all schools in China (implements a rigorous curriculum in all classes), because the new government policy calls for an end to the

tracking practices during the nine-year compulsory education period, but the practices are allowed in high schools.

Narrative Feedback on Content Validity

Members raised questions on the intended meaning for several items. Some of the questions could be easily answered by modifying the translation for certain words for clarification, for example, "schedule" in Item 26 (plans a schedule that enables quality instruction) needed to be clarified, and "program" in Item 51 (implements programs to help address community needs) became more clear when examples of such programs were given.

Other questions, however, were related with the contextual differences that have the potential to influence the construct validity of the framework and the assessment items. For Item 27 (coordinates efforts to improve instruction in all classes), several members thought it was unclear what "efforts" might be involved and what purpose such efforts served. For Item 28 (recruits teachers with the expertise to deliver instruction that maximizes student learning), members pointed out that most of the school principals do not have the hiring authority. For Item 71 (analyzes the influence of faculty evaluations on the rigor of the curriculum), members were concerned over two issues: principals have little decision-making authority over the curriculum content, let alone teachers; systemic accountability is still a relatively new concept.

When asked if there were core components and key processes that might be missing from the framework, six out of 12 members said none was missing, and six had something to add. For core components, "teacher professional development in learning

theories and new technology" was suggested by two members; for key processes, "motivating" was suggesting by three members, and "regulating" was suggested by two members. Overall, panel members did not believe that significant expansion of the framework was necessary, an indication of a general agreement that the framework largely reflect the leadership domain that addresses student learning.

The ratings on individual items brought up questions from the panel regarding possible contextual differences and their impact on the validity of the assessment. Items that show a negative gap between what is considered important and relevance could be interesting. It might be an indication that the members felt that too much emphasis might have been put on certain leadership behavior in practice, more than the importance that it deserves. For example, several members gave higher relevance ratings than for importance for Item 37 (plans programs and policies that promote discipline and order), noting that actions taken on discipline and order might stifle creativity. Items that emphasize rigor for curriculum and learning standards prompted comments from members on the direction of policy initiatives for education reform. "We are trying to reduce the 'rigor' of our content and to decrease the focus on achievement testing, aren't we?" one member asked. Items that are related to the key process of advocating, which emphasize the needs of all students especially those with special needs such as physical or learning disabilities in the English version, were translated to target students having difficulties in learning in the Chinese version. However, almost all members pointed out that advocating for the needs of students lagging behind academically had not been a priority and would be difficult to carry out. It might be especially difficulty for principals

of Chinese high schools where the practice of tracking still remains common and considered necessary to remain competitive for college entrance exams.

Summary

The existence of the "reality gap" and the variation in the gaps draw attention to the necessity of considering the multiple facets of content validity. Should the construct alignment be anchored on what it is now, or what it should be? Whose views are closer to the truth, the principals who work daily in the school setting or the researchers who are more intimately involved in standard-setting and professional development? In other words, how should the views of different experts be balanced?

The sizes of the "reality gap" vary among the six core components, and the perceptions on the gap are slightly different between the practitioner and the researcher groups. While the overall ratings indicated that the leadership behaviors of the VAL-ED core components and key processes are largely aligned with the view of Chinese experts on their importance to the success of schools and students, the alignment with the current practices of school principals were seemly weaker. Thus the significance of the findings on the "reality gap" is twofold: First, the VAL-ED framework and its leadership behavior items appear to be more content-valid when aligned with the perceived standards than with the current practice; second, the gap between what should be practiced and what are in practice points to an window of opportunity for future development of professional standards and training for school leaders.

Furthermore, the panel members called for attention on small or nonexistent "reality gaps" on some components and processes, which yielded interesting findings.

For example, the core component of Rigorous Curriculum shows the smallest reality gap while Connections with External Communities has the largest. Does this mean that Chinese principals have reached a high level of effectiveness in the areas with small gaps, or is the gap also an indication that this is an area that outside influences such as standardized curriculum at the national level leaves little room for principals to add their leadership value?

The expert-panel alignment analysis provided the first-round of feedback from both practitioners and researchers on the Learning-Centered Leadership Framework and the VAL-ED items. However, the panel is only comprised of four practitioners and eight researchers, and the input was limited to email communication. The significance of the rating, although preliminary, points out areas that need further investigation in the subsequent studies.

Study 2: Validation of Instrument Content

The validation of the instrument content featured cognitive interviews with two officials of a local bureau of education, two principals, and two teachers, in the Guangzhou area. 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.

Interviews with the school directors were conducted at the district offices during regular office hours. Interviews with the principals and teachers were conducted at the office space provided by South China Normal University after the school day. Each

participant was asked to assess a real principal: self assessment for the two principals, the

principal of the elementary school for the two teachers, and principals that they

supervised for the school directors. Each interview took about two hours. The interview

protocols were modeled after the process used in the U.S. validation work for the VAL-

ED (Porter et al., 2008). 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 ask how the item was perceived,

whether it was easily understood, and if there were any places that were ambiguous or did

not apply.

The focus of the interviews is to validate the content of the VAL-ED as an

instrument, including all elements such as directions for filling out the assessment,

sources of evidences for rating, the effectiveness scale, and the items themselves. In

addition, participants of the interviews also made suggestions on how to conduct the

assessment in the sampled schools for the third study.

Directions

On directions for the assessment, the participants thought the opening

statement about the purpose of the instrument itself was very clear. However, it

was recommended that additional elements be added: (a) a letter style opening

addressing the participant of the assessment, which is a standard format familiar to

principals and teachers of Chinese schools ; (b) explanation for why a U.S.

instrument is being tested in Chinese schools to provide context for the effort; (c)

clarification on the fact that there are three parties involved in the "360-degree"

assessment process: the principal, the teachers, and the supervisor of the principal;

and (d) emphasis on that the results of the assessment will be used strictly for the

research, not as an official evaluation for the principal. After the modification, the

Chinese version has the following opening statement (the original VAL-ED

statement is underlined):

> "Dear teacher (or principal, 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. <u>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.</u> 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."

Specific directions for how to fill out the assessment were considered clear

by the participants. The first line of direction in the VAL-ED states that "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."

This statement explains the stem that asks "how effective is the principal at

ensuring the school" in performing the 72 leadership behaviors. The

understanding of this leading phrase is key to grasping that the intent of the

assessment is on leadership behaviors that impact student learning through

working with the members of the school community. In other words, it needs to

be kept in mind that the assessed behavior is not limited to what is done by the principal him/herself. Participants of the interviews thought the statement was clear but raised this question: Are we considering everything happening inside of the school is impacted by the principal? Does this mean that in reality we are rating the school as a whole?

Sources of Evidence and Effectiveness Scale

Participants of the interview thought that the sources of evidence listed are self-explanatory and covered the most common sources in the school setting. Although one school director pointed out that by having "other" as a catch-all category, it gave too much leeway for people who don't want to select anything specific. Participants were asked to give examples of the sources of evidence to see if the examples were in the intended categories. For "school documents," examples were teaching assignment planning sheets, school reports that are either in print or online, and, very interestingly, public blogs that are kept by some principals; for "school projects or activities", examples were extracurricular programs in arts or sports, and civil service events. Several participants asked whether it might be more helpful to weigh the sources based on the extent to which they were used as evidence.

Participants also responded well to the effective rating scale. This type of rating was used by various surveys conducted throughout the school year and was familiar to the principals and teachers, according to the participants. During the rating of 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. Although for items related to

Connections with External Communities, several participants said because related

activities were so rare, frequency was a factor in their rating decisions. They also stressed

that they understood that it was the process involved in the principal's action that to be

rated. However, occasionally, the participants had to be reminded to refer to the stem of

the item when they asked questions on the possibility of the principal handling certain

tasks directly.

The biggest issue the participants had was with how sources of evidence and

effectiveness rating were sequenced for the assessment. The original version of the

VAL-ED that was piloted in the U.S. has the sources of evidence listed at the left and the

effectiveness rating on the right of each page. Directions on the first page ask

respondents to select sources of evidence first and mark the effectiveness rating for the

leadership behavior based next. Requiring respondents to select sources of evidence

prior to deciding on the rating is an important feature of the VAL-ED instrument to

enhance the validity of the results (Porter et al., 2008). However, all participants of the

cognitive interviews conducted in China unequivocally indicated that it was more natural

to the thinking process to decide on the rating first. "Yes the opinion is formed based on

evidence, but the feeling about how well it is done always comes to your mind first, then

you ask yourself which evidences formed this opinion," One respondent said. Another

respondent said: "Even if the evidences are listed on the left, I would still go to the right

side of the page to select the rating then come back to the evidences." On the rating scale,

the principals and teachers indicated that they would prefer a reversed order of 5-1,

listing outstandingly effective first, ineffective last. "In our cultural, what is good usually gets to be the first on a ranking list," One principal said.

The final Chinese version was modified to reflect the changes. The effectiveness rating was listed on the left and the sources of evidence on the right of each page. The rating scale was reversed to be from 5 to 1 on a Likert scale.

Item Contents

As a part of the iterative process of reaching construct equivalence between the original and the translated versions, four items that appeared unclear to the expert-panel participants were modified and then marked as needing more clarification during the cognitive interview sessions. For Item 27, "coordinates efforts to improve instruction in all classes", "not just focusing on fast-track classes" was added; for Item 51, "implements programs to help address community needs", a couple of examples of programs were put in parentheses; for Item 55, "promotes mechanisms for reaching families who are least comfortable at school", the example of migrant worker family was added; and for Item 71, "analyzes the influence of faculty evaluations on the rigor of the curriculum", the sentence was rewritten to further clarify that the intent of the leadership behavior is to make sure that faculty evaluation as one of the accountability measures would produce learning results. The four items were pointed out to the participants using probing questions such as "when you consider this item, what does the term . . . mean to you?" The answer of the participants was carefully compared with the original intent of the VAL-ED item. If a concept appeared to be abstract to the participant, examples might be

used for further clarification, and for the participant to make suggestions on possible changes to the wording.

Small verbiage changes were suggested by the participants as they went though the items to either make the sentences flow better or easier to understand. These changes were duly recorded. In addition to the linguistic suggestions, the participants provided a significant amount of information regarding how the items apply in urban Chinese schools during the cognitive interviews. Some of their responses were voluntary, some came about after probing. Using the core component of High Standards of Student Learning as an example, based on the questions raised by the expert panel regarding how the VAL-ED items were aligned with the directions of new educational reform initiatives, the interviewer checked if the respondent understood the intent of the item by probing: 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?

Their comments are presented in the order of first on the core components, then on the key processes of the VAL-ED.

The items for the first core component, High Standards of Student Performance, prompted interesting discussions. Respondents were asked: What do "standards" mean to you, policy mandates or internal organizational goals? How do you define "learning"? In your opinion, how is the level "high" for standards measured? What is high and what is low to you?

Each participant studied the definition for this core component and contemplated the translation for "student learning", which is the combination of two words, studying

and developing. 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 the latter was more of a priority for Chinese schools and students. 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 that 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. Participants also asked if there was an intended focus for this component in the U.S. Were both academic and social learning equally important, or the focus was more on academic achievement?

Participants gave relatively high ratings more consistently to the items for the second core component, Rigorous Curriculum, usually 4 or above when assessing a principal. The reason, according to the participants, was more related to the centralized curriculum standard-setting practice and the uniformity in textbooks and teaching guidelines. "Most teachers follow the teaching guide closely and there is little room to go off for something different," a teacher commented. According to the school directors, some schools were more innovative than the others in supplementing the standardized curriculum with school-based materials. One school director used the principal that he was assessing as an example for implementing informational technology courses that were developed by in-house faculty, which received national recognition.

For the 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 is used (VAL-ED Item 35), fall within the responsibility of one of the assistant principals or the instruction director. Participants believed that ratings were likely to be high thanks to the fact that 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 achieve high quality instruction is reached by working with administrators assigned to lead instructions and teachers' collaborations. However, quality of teaching varied greatly among teachers, according to the school directors and principals. 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 school districts had piloted school-based hiring practices, therefore items regarding teacher hiring might not be applicable.

Regarding the fourth core component, Culture of Learning and Professional Behaviors, participants believed that 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 through out 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 focused 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. Consequently, ratings on the items for this component appeared to be lower than for the others. "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 mentioned new initiatives at the district level such as the promotion of the "Three New Excellence" program that aimed at helping students to become excellent citizen in the society, excellent student at school, and excellent child at home. According to the participants, communications with families and the community often were generated organically from 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, were easy to understand. Systemic accountability as a term has a direct translation counterpart in Chinese. But this term was still relatively new to the education system and might not be familiar to all the teachers, commented by one of the teachers interviewed. For example, Item 71 was modified to reflect the expert-panel feedback. Because the intent of the item is to identity the influence of faculty evaluations on learning results, not necessarily just on curriculum content, the Chinese translation broadened the phrase "rigor of the curriculum" to "the improvement of teaching quality". Participants thought this item was now clear and applicable in the Chinese setting.

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 formulas and the principals had little control over the funds that were allocated for specific purposes. Support for building the learning culture financially (Item 41), for example, might be difficult, according to the interviewed participants.

Items related to the key process of advocating that describe leadership behaviors acting on behalf of the diverse needs of students were identified by the participants of the interviews as "more of a wish than practical reality" in Chinese urban schools. For example, on Item 7 and Item 8 that measure the effectiveness of the principal at ensuring

that the school attend to students with special needs, both school directors commented that while it might be the intention of the principals and teachers, it was difficult to implement with large class sizes (at least 50 students per class) and the pressure of standardized testing. Again, the phrase of students with special needs was translated to refer to students having difficulties in learning. For Item 19 (advocates 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 was not realistic to expect this leadership behavior at the school level with the word "ensuring", because of the standardized curriculum content and the achievement testing structure. Similar comments were made by the teachers on Item 31 (advocates for all students to regularly experience effective instruction) that students that 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 senior classes of middle and high schools.

Communicating as a key process was generally clear. However, when it is intersecting with the Core Component of Connections with External Community, the perceived weakness in working with families and the community at large by schools across the board seemed to be heightened by the key process of communicating.

The feedback on the processes of supporting, advocating and communicating was very relevant to the validity of the VAL-ED framework and items. Any item that is not applicable in the Chinese educational system such as Item 41 as it is written will need to be replaced. Items that are at the center of the reform debate regarding advocating for students of special needs and allocating administrative time to external connections may

not be suitable assessment indicators until the standards for principals are determined at the appropriate administrative level.

There were also some indications of possible "bleeding" between the two dimensions, core components and key processes, of the framework. This type of "bleeding" seemed to be more noticeable by the participants between Connections with External Communities and communicating, and between Systemic Performance Accountability and monitoring. Specifically, when communicating with parents was used in other core components, participants felt that the items were somewhat redundant. For example, Item 69 (discusses progress toward meeting school goals with parents), were mentioned by several participants that might be redundant with items that they had rated earlier. Similarly, several participants thought Item 11 (monitors student learning against high standards of achievement) was not very different from Item 65 (allocates time to evaluate student learning).

Suggestions for Assessment Data Collection

Upon completion of the VAL-ED instrument, each participant of the cognitive interview was consulted for their input on the process of conducting the next study, where the VAL-ED rating scales would be given to selected principals, teachers in these principals' schools, and supervisors of the principals.

From talking with the participants, two sampling-related issues were discovered. First, which supervisor should rate the principal on his or her "learning-related" performance; second, which top school administrator is the true principal that would be evaluated. These may seem to be odd issues in the U.S. context. In the U.S., the

personnel structure for managing principals is relatively clear. The inventory of the

practice of principal assessment in urban U.S. districts (Goldring et al., 2008) found that

the responsibility of evaluating school principals is usually carried out by the assistant

superintendent in charge of curriculum and instruction. However, according to the school

directors and principals interviewed, principals in urban Chinese schools report to

multiple administrative offices that have jurisdictions over an array of functions of the

schools. When asked who would be the most appropriate person to rate the principal's

performance regarding student learning, neither school directors could pinpoint a

department or person.

      The management structure at the principal level has a feature that remained from

the central party control era before the 1980's. Typically each Chinese school has two

top administrators, one is usually called the headmaster; the other is called the secretary

general. While the two administrators have the same civil service rankings, they have

different responsibilities. The secretary general is in charge of political functions such as

membership of the teacher's union, communist party, youth and the youth league. The

role of the headmaster is very similar to that of a school principal in the U.S. Both types

of leaders participate in university-run principal training programs. It is common when

both are called the principal as a generic term. Even more confusing, some principals

carry the responsibilities of both and the business card would list both titles.

      Upon discussions that clarified the roles and responsibilities of the school district

offices, the decision was made that the school directors would be sampled as supervisors

for the principals, because through reports and input from the district administrative

offices, the director is the one that has the most comprehensive knowledge of the

principal's performance. It was also decided that the secretary generals were not appropriate subjects for the VAL-ED assessment because they usually do not have direct responsibilities over curriculum and instruction, which are important elements of the learning-centered assessment.

Study 3: Reliability and Validity of the VAL-ED Scores

The Pilot Test Sample

The validity and reliability examination of VAL-ED score results was carried out for 19 schools. The final sample included 18 principals, 1165 teachers, and 6 school directors that are supervisors of the 18 principals. The final version of the VAL-ED assessment in Chinese was used in the data collection. Twenty schools were initially selected according to the sampling plan. But one school dropped out because its principal was on leave unexpectedly and assessment arrangements could not be made.

The assessment data collection effort took the span of five working dates. Among the 19 schools, the assessment were administered in group settings and collected at the end of the session in 13 schools. The introduction of the research team was given by the assessed principal at every group session, followed by the explanation of the research purpose and instruction for completing the assessment by the research staff. This process was helpful in raising the teachers' comfort level in rating their principal, a practice that is not very common in China schools. In the other six schools, assessments were distributed either by the principals or at the end of faculty meetings by the research staff, and collected a few days after. Among the 19 principal self-assessments to be collected, 18 were obtained. One principal decided at the last minute that he was uncomfortable

with evaluating himself and withdrew his participation. However, he had no problem

with being assessed by the teachers in his school.

Missing Data

Data was examined in terms of missing pattern and missing frequency. Patterns

that show particularly high rates of missing data in certain schools or certain clusters of

items may be indications of threats to validity for the instrument. While acceptable

missing data rates are usually determined case-by-case depending upon the research

scope, missing data at below 10% is often used as a general cutoff (Porter et al., 2008).

Missing data occurred but the problem appears to be relatively small in size. For

the teacher assessment, 15% of the observations miss more than 10% of the effective

rating scores (Table 8). The missing data rates for principal and supervisor assessments

were lower than 10% per observation across the board.

Table 8: Teacher Assessment Missing Data by Observation

Pct of Effectiveness Ratings Missing	# of Observations	Pct of Total Observations
<10%	986	84.64
10-20%	54	4.63
20-30%	23	2.06
30-40%	10	0.86
>40%	92	7.81

An examination of the scores revealed no apparent systemic patterns of missing

data by schools. Table 9 shows the average number of missing effectiveness ratings by

school, and Figure 6 illustrates a normal distribution of the school averages for missing

data points for the effective ratings. A further examination of the missing ratings for each

school (Figure 7) shows very similar patterns of a positively skewed distribution,

indicating that most of the observations have less than 10% of missing data points.

Table 9: Average Missing Effectiveness Ratings by School

School ID	Average Missing Ratings	Standard Deviation	# of Returned Assessments	Return Rate
1	8.29	18.99	55	100.00%
2	8.70	18.59	83	95.40%
3	6.93	19.01	67	98.53%
4	9.63	23.21	59	90.77%
5	10.93	14.05	27	100.00%
6	5.82	18.83	149	97.39%
7	9.37	23.01	120	99.17%
8	7.74	19.37	35	106.06%
9	3.63	10.78	35	97.22%
10	0.77	1.94	35	97.22%
11	4.75	16.48	56	93.33%
12	6.58	19.05	40	97.56%
13	11.45	22.08	47	100.00%
14	4.11	15.24	45	88.24%
15	5.28	14.47	155	100.00%
16	2.97	12.64	33	100.00%
17	0.80	3.64	30	100.00%
18	2.33	5.04	43	97.73%
19	12.75	24.36	51	98.08%
Total	6.75	18.08	1165	97.40%
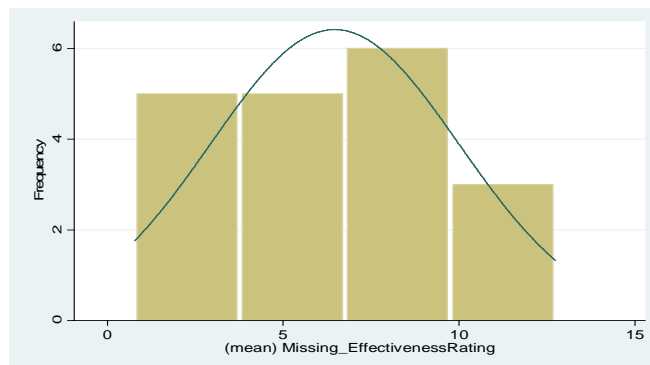
Figure 6: Distribution of Average Number of Missing Ratings by School
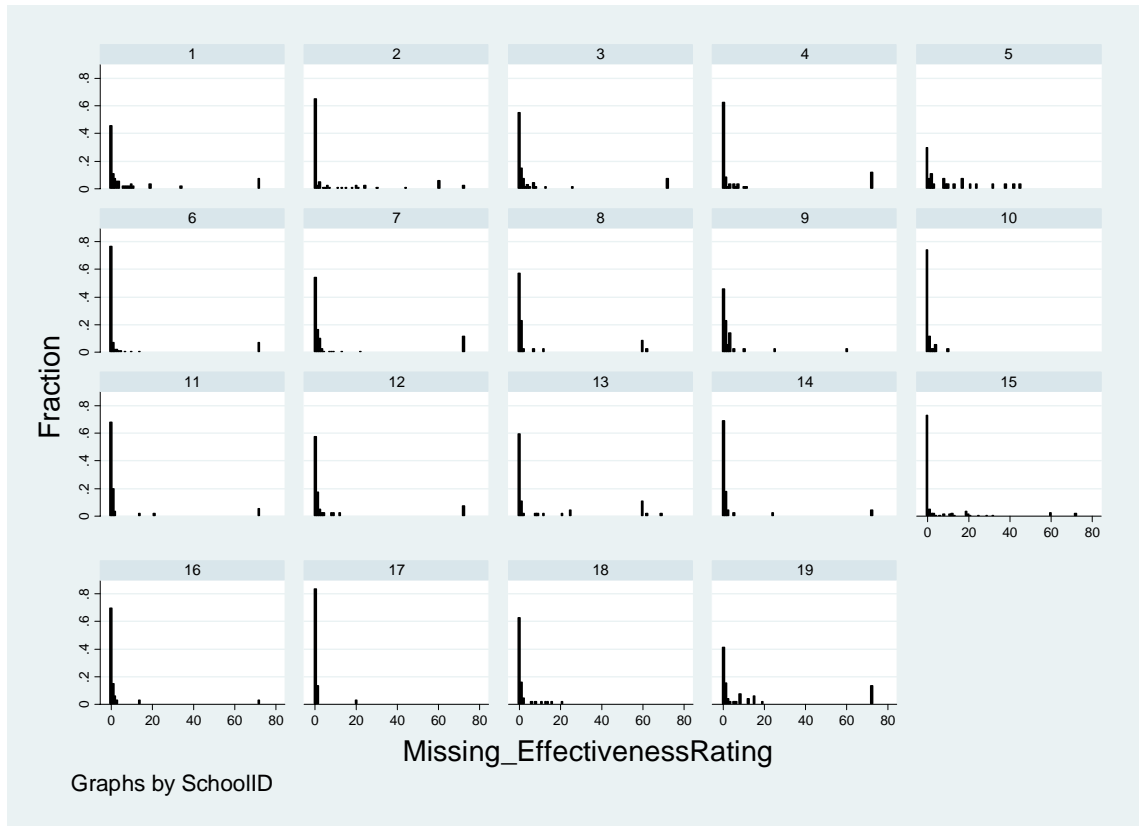
Figure 7: Distribution of Missing Data <u>within</u> School

Table 10: Teacher Assessment Missing Data Summary by Item

Component 1		Component 2		Component 3		Component 4		Component 5		Component 6	
Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing
1	7.98%	13	8.76%	25	9.70%	37	5.75%	49	12.19%	61	8.07%
2	7.81%	14	9.79%	26	8.15%	38	6.87%	50	10.30%	62	10.04%
3	8.24%	15	8.93%	27	8.93%	39	6.27%	51	13.65%	63	10.04%
4	8.24%	16	9.44%	28	9.01%	40	10.13%	52	8.24%	64	9.79%
5	8.67%	17	8.67%	29	8.93%	41	8.58%	53	11.59%	65	10.73%
6	8.58%	18	8.33%	30	7.98%	42	6.01%	54	12.45%	66	9.87%
7	8.50%	19	10.21%	31	8.33%	43	7.55%	55	13.05%	67	9.36%
8	8.58%	20	8.41%	32	9.01%	44	8.07%	56	13.82%	68	8.58%
9	8.41%	21	8.67%	33	7.90%	45	8.76%	57	10.47%	69	11.33%
10	11.24%	22	9.36%	34	9.36%	46	6.87%	58	10.90%	70	9.79%
11	9.53%	23	9.10%	35	9.87%	47	7.98%	59	12.36%	71	10.39%
12	9.44%	24	10.39%	36	9.61%	48	7.81%	60	14.94%	72	10.30%

The missing scores were also checked at the item level (Table 10). High

percentages of missing data are threats to construct validity and could be caused by

various reasons: poor translation, sensitive or uncomfortable subject, inapplicable

scenarios, or confusion over the question. The missing-data percentages for the 72 items

hover around the average of 9%, with a standard deviation of 2%, ranging from the

lowest (Item 42) at missing 6% to the highest (Item 60) at missing 15%. The distribution

of the item-level missing-data percentages is quite normal (Figure 8).
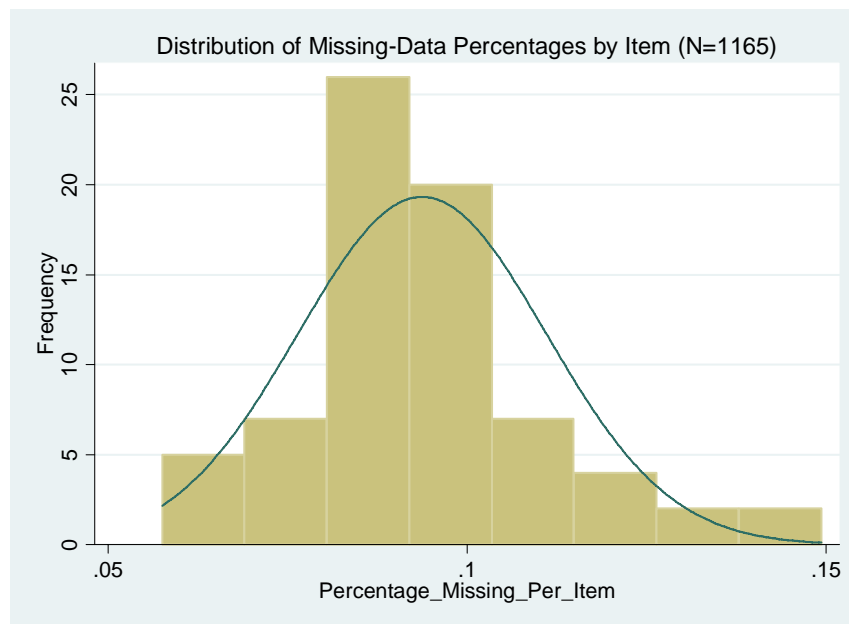
Figure 8: Distribution of Missing-Data Percentages by Item

From eyeballing Table 10, we see that items with higher percentages of missing

data appear more in Component 5, Connections to External Communities (Item 49-60)

and Component 6, Systemic Performance Accountability (Item 61-72). To see where the

missing data occur most, the missing scores were subsequently sorted by core

components and key processes. Table 11 shows that percent of missing data for each

scale hovers around 9%. Connections to External Communities (Component 5) and

Systemic Performance Accountability (Component 6) indeed stand out as having the

largest proportions of missing data at 12% and 9.86% respectively. Items of High

Standards have the smallest mean percentage of missing data at 8.77%. Among the key

processes, the largest missing-data percentage averages are for monitoring at 10.14% and

advocating at 9.46%. It is interesting that the subscales with relatively higher rates of

missing data contain items that were considered as less fitting or needing clarification by

the expert panel and the cognitive interview participants.

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. Also from Table 11, we see

that this percent is very low for all the VAL-ED scales at 1-2% of the total observations,

The only exception is the core component of Connections to External Communities,

which is at 4.5%.

Table 11: Teacher Assessment Missing Data Summary
– by Core Components and Key Processes

	Missing	Don't Know
Core Components		
High Standards	8.77%	0.90%
Rigorous Curriculum	9.17%	1.61%
Quality Instruction	8.90%	1.07%
Culture of Learning and Prof. Behaviors	7.55%	1.39%
Connections to External Communities	12.00%	4.50%
Systemic Performance Accountability	9.86%	2.12%
Key Processes		
Planning	8.78%	1.64%
Implementing	9.24%	1.58%
Supporting	9.20%	1.65%
Advocating	9.46%	1.92%
Communicating	9.42%	2.04%
Monitoring	10.14%	2.75%

Factor Analysis

The exploratory factor analysis results based on the teacher effectiveness rating are shown in Table 12. Recall that the purpose of the EFA is to analyze the common variance in the variables and it is a data reduction technique that reduces the number of dimensions that are needed to explain the common variance. A cutoff of .40 was used to interpret solution from the orthogonal rotation. Items that loaded high on one factor and relatively low on all the others were marked in Table 13, whereas items with low loadings and/or dual-factor loadings were not listed.

Among the 72 VAL-ED items, 59 (82%) of them have factor loadings larger than .40 and 13 do not. The 59 items load on to seven factors, three of which have Eigen values larger than 1.0, accounting for 87.7% of the cumulative (total) variance. The other four factors account for another 6.0% of the total variance. Together, the seven factors account for 93.7% of the total variance.

The EFA results show a factor structure that reflects the theoretical framework of six core components only to a limited extent. The most noteworthy discrepancy is the lack of discriminatory power of the items to reflect the six-component framework. The common variances of the items are concentrated on three main factors. Among the 59 items that have significant factor loadings, 26 of them are concentrated on one factor explaining 81.9% of the total variance. There are only two other factors that have 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 9),

when the Eigen value gets close to 1.0, the plot level levels off to a linear decreased

pattern.

Table 12: Structure Matrix for Exploratory Factor Analysis

	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5		Factor 6		Factor 7	
	No.	1	No.	2	No.	3	No.	4	No.	5	No.	6	No.	7
	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												
E.V.*	42.56		1.80		1.22		0.93		0.86		0.69		.63	
C.V.*	81.92%		85.40%		87.74%		89.53%		91.18%		92.51%		93.71%	

*E.V. Eigen Value; C.V. Cumulative Variance
NOTE: Extraction method – principal axis factoring – was used. For clarity, only values equal to or greater than .40 are provided. Among the 72, 13 items have loadings lower than .40 and are not listed in the table. They are items 6, 8, 9, 18, 20, 21, 22, 23, 24, 37, 46, 61, 62.
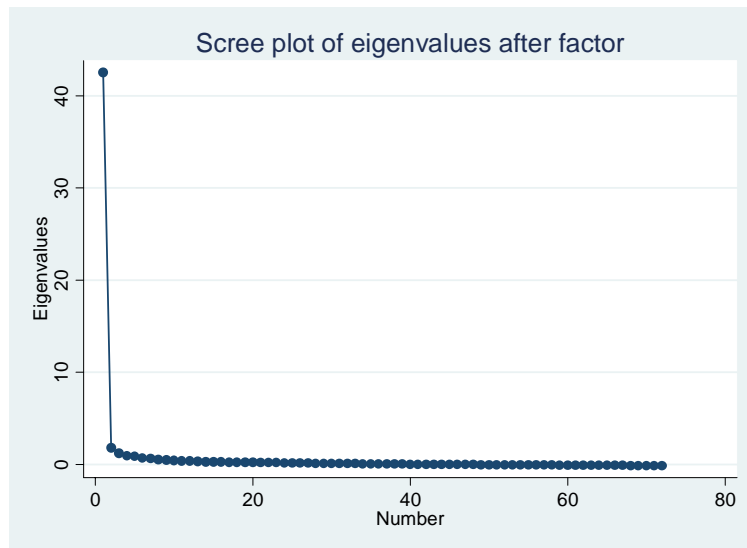
Figure 9: Scree Plot of Eigen Values for Factor Determination

To better understand how the factor structure based on the empirical data fit the VAL-ED framework factor loadings larger than .40 were retained and plotted into the two-dimension matrix for core components and key processes based on their factor clusters despite the cluster Eigen value.

Core Components	Key Processes											
	P		I		S		A		C		M	
High Standards for Student Learning	4	4	4	4	4		3			1	3	3
Rigorous Curriculum	3	3	3	3	3		3					
Quality Instruction	3	2	2	2	2	2	2	1	2	6	6	6
Culture of Learning and Professional Behaviors		7	7	7	1	2	1	2	1		1	1
Connections to External Communities	1	1	1	1	1	1	1	1	1	1	1	1
Systemic Performance Accountability			1	1	1	1	1	1	1	5	5	5

Figure 10: Illustration of Factor Structure in the Current LCL Framework

The results are shown in color in Figure 10.   The first four core components are more distinctively represented by the factor structure. Five items of High Standards for Student Learning cluster around Factor 4, six items of Rigorous Curriculum around Factor 3, seven items around Factor 2, and three around Factor 7. The clustering of items around the principal factors is by no means clean-cut and some "bleeding" exists among the four components.

The items clustering around Factor 1, the largest factor, are 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 does 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 are also marked by Factor 1. This very well may be an indication of the bleeding of the last two key processes with the last two core components, 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 1165 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 are sufficiently distinctive from one and another to embody valid constructs individually and combined for learning-centered leadership. Important discrepancy between the theoretical framework of the VAL-ED and findings do 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% of covariance. Furthermore, the factor structure is not clear. The item clustering patterns appear to be scattered across components and processes, and too many items clustered around the last two core components. The findings indicate that modification and refinement to the items and even possibly to the framework may be necessary to reach a better level of fit of actual factor structure and the theoretical framework.

On a more technical note, the findings of the orthogonal rotation are supported by the alternative oblique rotation results, which also yielded three main factors with Eigen values larger than 1.0 accounting for nearly 90% of the common variances. The factor structure as defined by clustered items using the oblique rotation was even less clear than that by the orthogonal rotation, however. "Usually when multiple factors are extracted, reasonable simple structure is realized with varimax rotation. And if these factors are interpretable, the correct rotation method has typically been identified" (Thompson, 2004, p.48). While the orthogonal rotation results are used for the factor analysis for this study, because the latent constructs have noteworthy correlations theoretically (Murphy et al., 2006; Porter et al., 2008), an oblique rotation (e.g., promax) may be suitable in future analyses to obtain factor scores when the expected factor structure of the modified frame and instrument items is better justified in the new cultural setting.

Reliability Estimates

The internal consistency of the VAL-ED items was first examined with the teacher VAL-ED scores for the Cronbach's alpha values on the subscales of core

components and key processes. The results are presented in Table 13. Based on the

teacher assessment effectiveness ratings, all subscales exhibit excellent internal

consistency with Cronbach's alphas larger than .90 and most of them at about .95. The

alpha value for the 72-item overall scale is .96.

Standard errors of measurement (SEM) were calculated for the mean scores of

components, processes and the full-scale average. The SEM is the standard deviation of

a hypothetically infinite number of obtained scores around the person's true score. SEM

allows us to estimate the degree to which a test provides inaccurate readings. If

reliability = 1, then SEM = 0 (no measurement error). The smaller the SEM, the more

certain we can be about the accuracy with which an attribute is measured (Elliott, 2008).

In this case, we see (Table 13) that because the reliability measured by Cronbach's alphas

is high at the sub-scale and full-scale levels, the standard errors of measure are very low,

and the confidence intervals for the mean scores have relatively small ranges, providing

further evidence for strong internal reliability and the accuracy of the assessment results.

Table 13: VAL-ED Scale Reliability

Scale	Cronbach's Alpha	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 Prof. Behaviors	0.951	4.34	0.62	0.14	4.11	4.57
Connections to Ext. 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 14: Between-Subscale Correlations of Core Components

	CC1	CC2	CC3	CC4	CC5	CC6
Principal VAL-ED (N=18)						
CC1	1.00					
CC2	0.71	1.00				
CC3	0.76	0.71	1.00			
CC4	0.38	0.59	0.34	1.00		
CC5	0.44	0.44	0.51	0.03	1.00	
CC6	0.70	0.67	0.64	0.44	0.42	1.00
Supervisor VAL-ED (N=18)						
CC1	1.00					
CC2	0.82	1.00				
CC3	0.88	0.87	1.00			
CC4	0.75	0.87	0.89	1.00		
CC5	0.89	0.88	0.88	0.88	1.00	
CC6	0.82	0.86	0.74	0.77	0.79	1.00
Teacher VAL-ED (N=1165)						
CC1	1.00					
CC2	0.97	1.00				
CC3	0.96	0.95	1.00			
CC4	0.98	0.97	0.97	1.00		
CC5	0.96	0.92	0.96	0.95	1.00	
CC6	0.97	0.95	0.97	0.98	0.97	1.00

Table 15: Between-Subscale Correlations of Key Processes

	KP1	KP2	KP3	KP4	KP5	KP6
Principal Assessment (N=18)						
KP1	1.00					
KP2	0.47	1.00				
KP3	0.14	0.65	1.00			
KP4	0.34	0.87	0.56	1.00		
KP5	0.46	0.81	0.46	0.80	1.00	
KP6	0.42	0.86	0.64	0.85	0.86	1.00
Supervisor Assessment (N=18)						
KP1	1.00					
KP2	0.93	1.00				
KP3	0.83	0.82	1.00			
KP4	0.89	0.88	0.79	1.00		
KP5	0.90	0.83	0.77	0.88	1.00	
KP6	0.93	0.91	0.85	0.94	0.88	1.00
Teacher Assessment (N=1165)						
KP1	1.00					
KP2	0.98	1.00				
KP3	0.96	0.97	1.00			
KP4	0.98	0.99	0.95	1.00		
KP5	0.96	0.98	0.95	0.97	1.00	
KP6	0.96	0.98	0.95	0.97	0.98	1.00

The internal consistency of the assessment instrument was also checked by obtaining between-subscale correlations. Correlations among the core components and among the key processes were obtained separately for the principal, supervisor and teacher assessments (Table 14 and Table 15). The between-subscale correlations are to show whether the mean scores given to the subscales are consistent. For example, we want to see when principals are rating themselves, how the ratings for the six core components co vary. The results show that in comparison, the teacher assessment scores show much higher correlations for both core components and key processes than the other two parties. Principal assessment scores show the lowest correlations. For example, within the principal assessment, Connections with External Communities (Component 5) has low correlations with all other core components and it is as low as .03 with Culture of Learning and Professional Behaviors (Table 14); planning has a .14 correlation with supporting and a correlation of .34 with advocating (Table 15).

The inter-rater correlation results are shown in Table 16. Correlations among the ratings of three parties, teachers, principals, and supervisors, were obtained for the 18 schools. The unit of analysis is the school-level mean score on the principal. The three sets of data include the averages of teacher ratings for the core components, key processes, and the instrument mean score, and the same categories of ratings by the principals and the supervisors for each school. Correlations are very low between teachers and principals. The correlation is .15 for the total mean score, and below .15 for all the components and processes. Correlations are even slightly negative for three components and three processes. Correlations between teacher and supervisor scores are similar to those between teachers and principals in that the mean score correlation is

minimal (.00), and there are some scales with negative but very small correlations. However, correlations between teachers and supervisors on two components, High Standards for Student Learning and Performance Accountability, and one process, communicating, have relatively higher correlations at above .20. As for correlations between principal and supervisor scores, the mean score correlation is also at a low .04, and there are three components and two processes with small, negative correlations. However, three correlations are higher than .30: High Standards for Student Learning (.37), Quality Instruction (.39), and Communicating (.33), indicating that the principals and supervisors appear 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 examines the inter-rater reliability of the assessment when it is used by respondents in the same role. The results are presented in the last column of Table 16. The correlations between two randomly selected teachers in each school rating the same principal are much higher than the teacher-principal, principal-supervisor, and teacher-supervisor correlations. The correlation of the total mean score is an encouraging .37. A majority of the sub-scale correlations for the components and processes are in the moderate level of between .30 and .59. The inter-rater reliability results indicate that teachers, principals, and supervisors tend to rate the effectiveness of the principal differently, and teachers within the same schools tend to have more agreement on how the principal performs in their school.

Table 16: VAL-ED Inter-Rater School-Level Mean Score Correlations (N=18)

	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

The inter-rater consistency was also measured using the Cohen's kappa coefficient, an alternative statistical measure of inter-rater agreement. The Cohen's kappa is generally thought to be a more robust measure than simple percent agreement calculation since κ takes into account the agreement occurring by chance. Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories. If the raters are in complete agreement then $\kappa = 1$; if there is no agreement among the raters (other than what would be expected by chance) then $\kappa \leq 0$ (Cohen, 1960). Note that Cohen's kappa measures agreement between two raters only. Landis and Koch (1977) gave a table for interpreting κ values where $\kappa < 0$ indicates no agreement, κ at $0.00 - 0.20$ shows slight agreement, and κ at $0.20 - 0.40$ indicates fair agreement. The table is however *by no means* universally accepted; Landis and Koch supplied no evidence to support it, basing it instead on personal opinion. The kappa will be higher when there are fewer categories.

Total mean scores of the VAL-ED by teachers, supervisors and principals were converted into four discrete categories (<3.0, 3.0 – 3.5, 3.5 – 4.0, >4.0) to cover the span of the skewed distribution of the scores. The agreements between raters were then measured by Cohen's kappa. The results are presented in Table 17. We see that the kappa coefficients are still very low between teachers and principals (.04), and between teachers and supervisors (.08). The kappa coefficient between principals and supervisors is slightly higher at .21, a fair agreement according to Landis and Koch (1977). The kappa measure further confirms the findings of low inter-rater agreement among the three parties indicating that teachers, principals and supervisors arrived at different final scores when they rated the same school principals.

Table 17: VAL-ED Inter-Rater Consistency by Kappa Coefficient (N=18)

	Kappa	Agreement	Expected Agreement	Std. Err.	Z	Prob>Z
Teacher/Principal	0.04	38.20%	35.80%	0.02	1.93	0.03**
Principal/Supervisor	0.21	52.53%	40.01%	0.02	9.81	0.00***
Teacher/Supervisor	0.08	44.21%	39.22%	0.02	3.90	0.00***

To delve deeper into possible reasons of the lack of inter-rater correlations for the core components, key processes, and the full-scale mean scores, the assessment scores were further examined first by the VAL-ED scale and then by school.

Table 18: Summary of Core Component Ratings

Core Components		Principals n=18	Supervisors n=18	Teachers N=1165
High Standards	Mean	4.07	4.50	4.29
	SD	0.46	0.45	0.64
	Max	4.83	5.00	5.00
	Min	3.33	3.58	1.50
Rigorous Curriculum	Mean	4.04	4.39	4.22
	SD	0.37	0.42	0.65
	Max	4.58	5.00	5.00
	Min	3.17	3.75	1.33
Quality Instruction	Mean	4.17	4.61	4.34
	SD	0.39	0.39	0.61
	Max	5.00	5.00	5.00
	Min	3.58	4.00	1.42
Culture of Learning and Prof. B.	Mean	4.08	4.53	4.34
	SD	0.48	0.48	0.62
	Max	4.83	5.00	5.00
	Min	2.92	3.75	2.08
Connections to Ext. Communities	Mean	3.75	4.29	4.17
	SD	0.50	0.57	0.74
	Max	4.75	5.00	5.00
	Min	2.50	3.08	1.42
Systemic Performance Accountability	Mean	3.81	4.44	4.31
	SD	0.78	0.50	0.66
	Max	4.92	5.00	5.00
	Min	2.17	3.58	1.92
Total Mean Rating	Mean	3.94	4.45	4.34
	SD	0.38	0.42	0.58
	Max	4.68	5.00	5.00
	Min	3.32	3.86	2.22

Table 18 and Table 19 list the mean scores of the core components and key processes from the principals, supervisors, and the teachers. As illustrated in Figure 11, the teacher VAL-ED scores are negatively skewed with a significant portion of the ratings in the range between 4 and 5, so are the supervisor assessment scores. The principal VAL-ED scores appear to have a more normal distribution.

Table 19: Summary of Key Process Ratings

| Key Processes		Principals N=18	Supervisors n=18	Teachers N=1165
Planning	Mean	3.93	4.54	4.33
	SD	0.71	0.42	0.58
	Max	4.83	5.00	5.00
	Min	2.08	3.75	2.17
Implementing	Mean	4.06	4.55	4.29
	SD	0.40	0.43	0.64
	Max	4.58	5.00	5.00
	Min	3.33	3.92	1.33
Supporting	Mean	4.07	4.51	4.29
	SD	0.52	0.33	0.63
	Max	4.67	5.00	5.00
	Min	2.50	4.08	1.67
Advocating	Mean	3.94	4.52	4.28
	SD	0.40	0.42	0.63
	Max	4.58	5.00	5.00
	Min	3.33	3.75	1.83
Communicating	Mean	4.07	4.52	4.30
	SD	0.48	0.45	0.65
	Max	5.00	5.00	5.00
	Min	3.33	3.75	1.58
Monitoring	Mean	3.84	4.38	4.22
	SD	0.38	0.59	0.66
	Max	4.50	5.00	5.00
	Min	2.92	3.42	1.67
Total Mean Rating	Mean	3.94	4.45	4.34
	SD	0.38	0.42	0.58
	Max	4.68	5.00	5.00
	Min	3.32	3.86	2.22
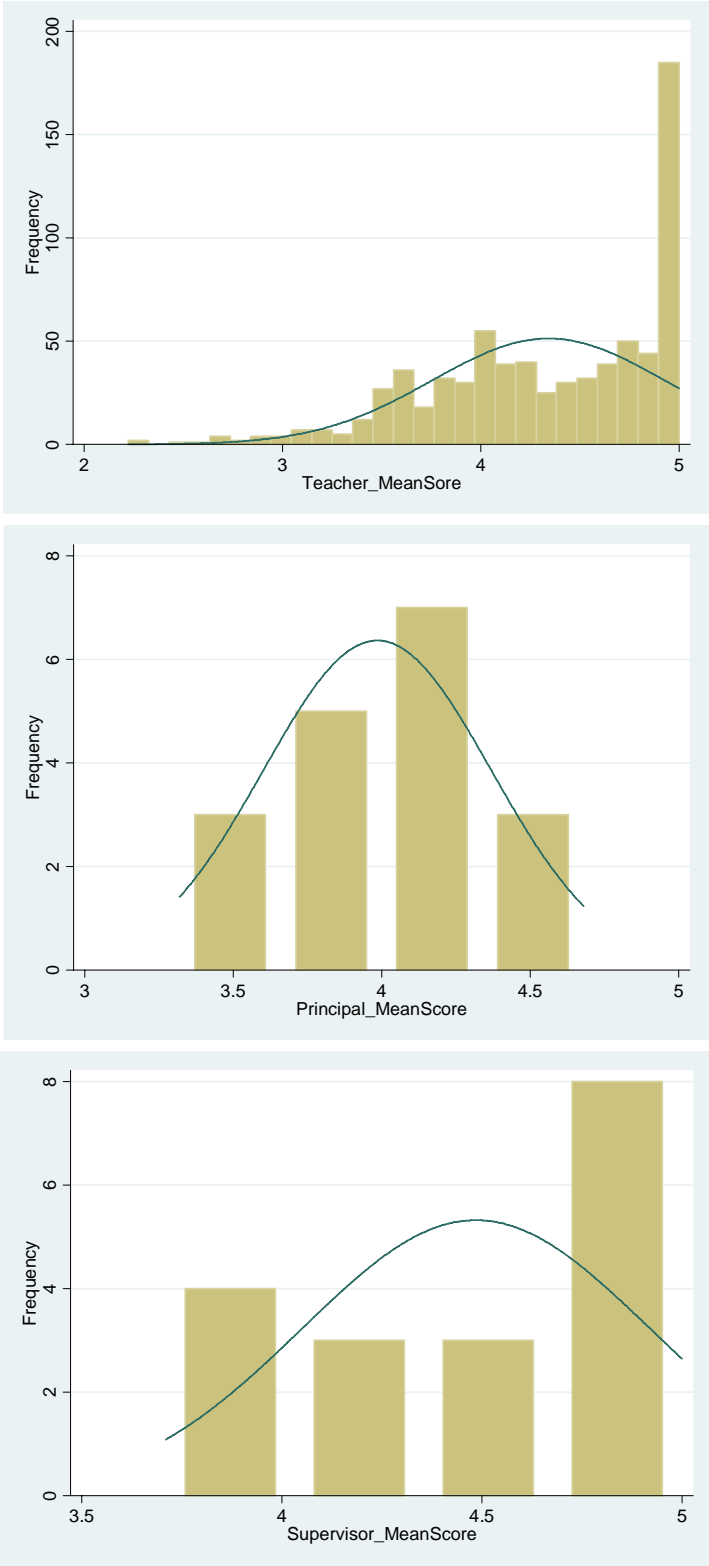
Figure 11: Distribution of VAL-ED Teacher Effectiveness Ratings

The descriptive summary in Table 18 and Table 19 shows that the mean scores of the principal, supervisor, and teacher assessments vary by core component and key process. The variations are further illustrated in Figure 12 and Figure 13. For the core components (Figure 12), the average ratings from the principals are consistently lower than ratings given by their supervisors and teachers. Supervisors gave the highest ratings in comparison, with the teacher ratings in the middle for all six of the core components. For example, while all parties gave the lowest average ratings to Connections to External Community, the principals gave an average of 3.75, the supervisor average was 4.29 and the teacher average was 4.17. The largest discrepancy between the principal self-ratings and the ratings of their supervisors and teachers is for Systemic Performance Accountability, with the principal average at 3.81, the supervisor's at 4.43, and the teachers at 4.31. For the key processes (Figure 13), again the principal ratings are the lowest, next to the teacher ratings then the supervisor ratings. For advocating, an area considered potentially weak due to lack of attention paid to students lagging behind academically or with diverse needs, the principal rating is 3.94, the supervisor rating is 4.52, and the teacher rating is 4.28.
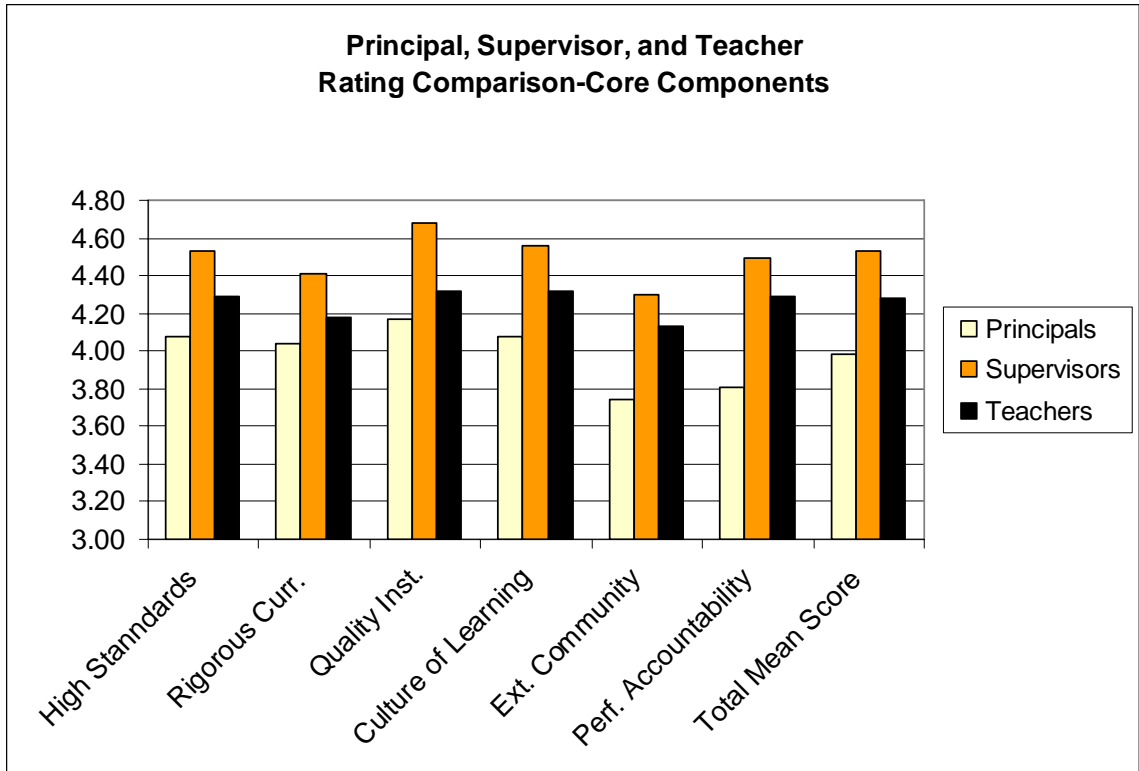
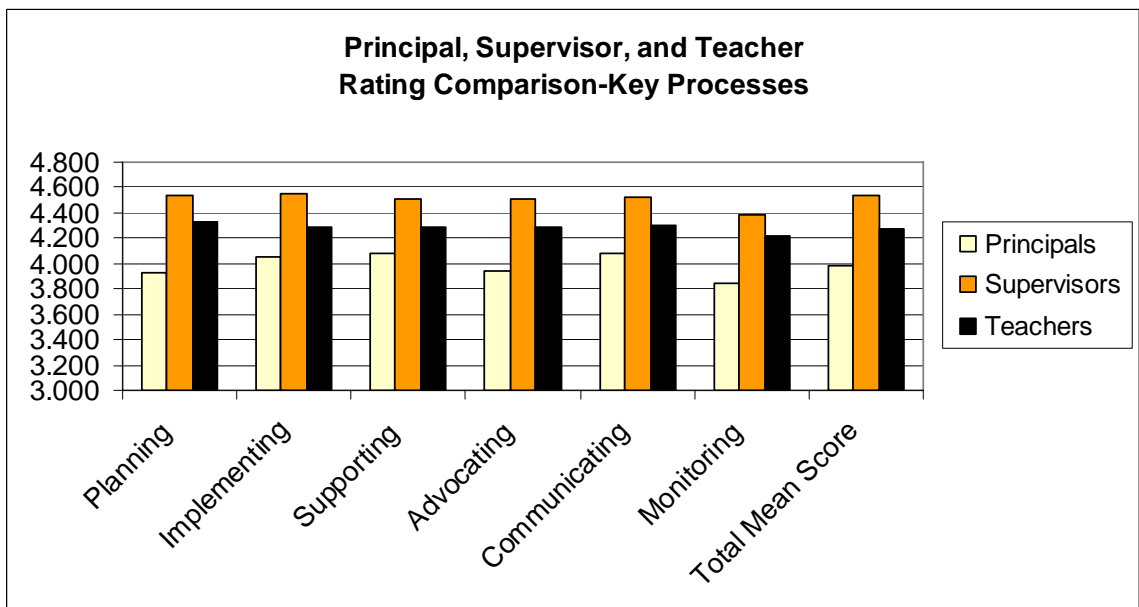Figure 12: Inter-Rater Variation in Core Components (N=18)

Figure 13: Inter-Rater Variation in Key Processes (N=18)

In addition to the different views of the three parties regarding the core components and key processes, principal, supervisor and teacher ratings also vary greatly by school. In other words, the variations among the three parties are different depending on which principal was assessed. Figure 14 demonstrates 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. Principal ratings range from 3.32 to 4.68, supervisor ratings range from 3.71 to 5.00, and teacher ratings range from 3.85 to 4.75.
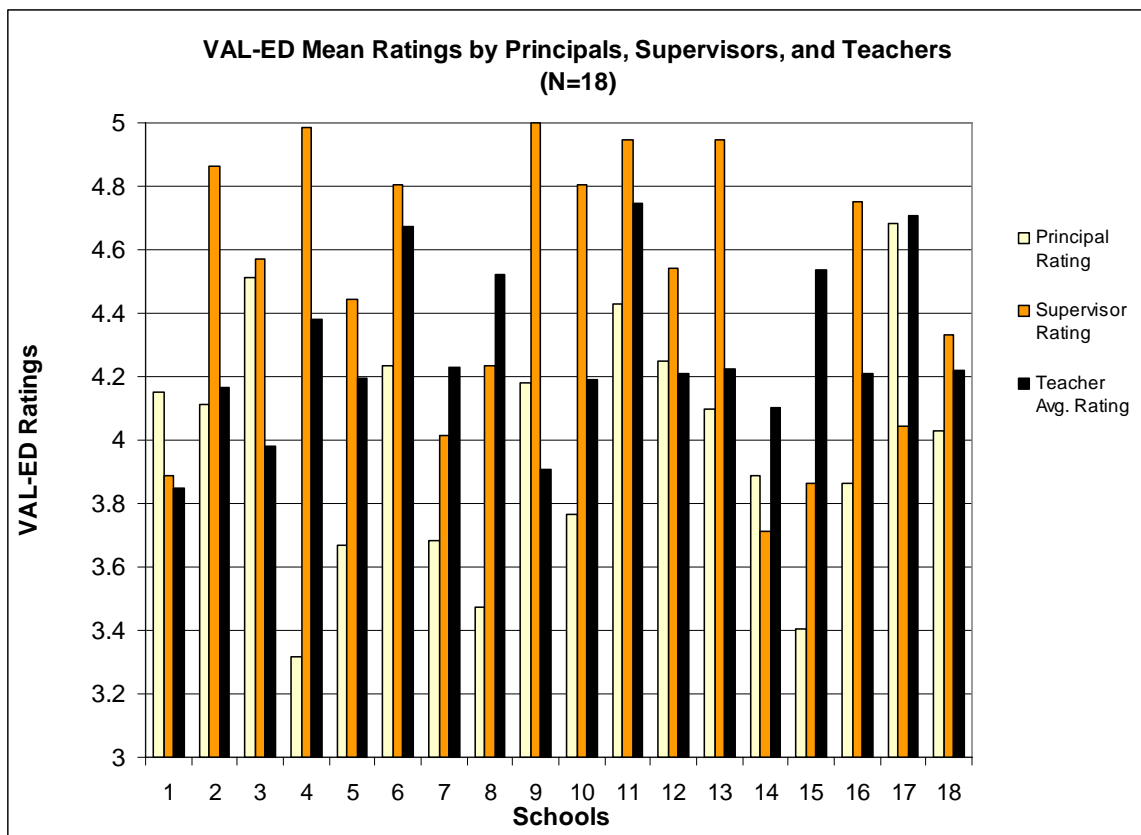

Figure 14: Inter-Rater Variation by School

Figure 15 and Figure 16 compare the rating patterns of the three parties of a school with *low* teacher ratings and a school with *high* teacher ratings on the core components. We see from Figure 14 that School 1 (N=55) has the lowest teacher ratings compared with the other schools also has the second lowest mean score from the supervisor. However, the principal self-ratings are much higher. We see that the supervisor rating is consistent with the teacher ratings on five of the six core components (except for Connections to External Communities), which are lower than the principal's, and the rating levels are very similar. It appears that in this case, the principal had a better self image of her/his leadership effectiveness, which was not quite the view of the supervisor and the teachers. In comparison, School 11 (Figure 16) that has the highest teacher rating (N=56) demonstrates a very different rating pattern. The principal self-ratings are lower than the supervisor's and the teachers' for all six core components. More importantly, the ratings of the teachers and the supervisor are high across the board and are similar across the six components and across the raters. The principal's self ratings do not differ very much from the supervisor and teachers. The only exception here is the principal self-rating for Connections to External Community was a low 3.42, echoing the sentiment of other practitioners toward the reality of this matter. It appears that in this school, the supervisor and the teachers agreed that the principal had an overall high effectiveness in leadership, and the principal's self-appraisal was consistent with the supervisor and the teachers.
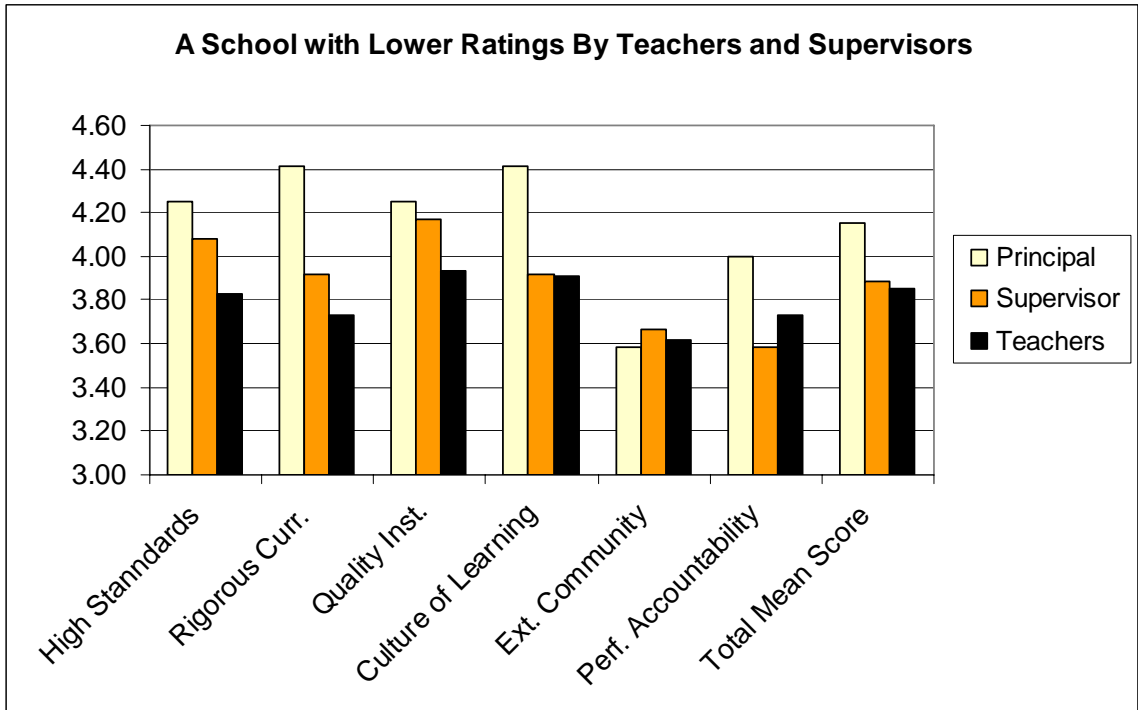
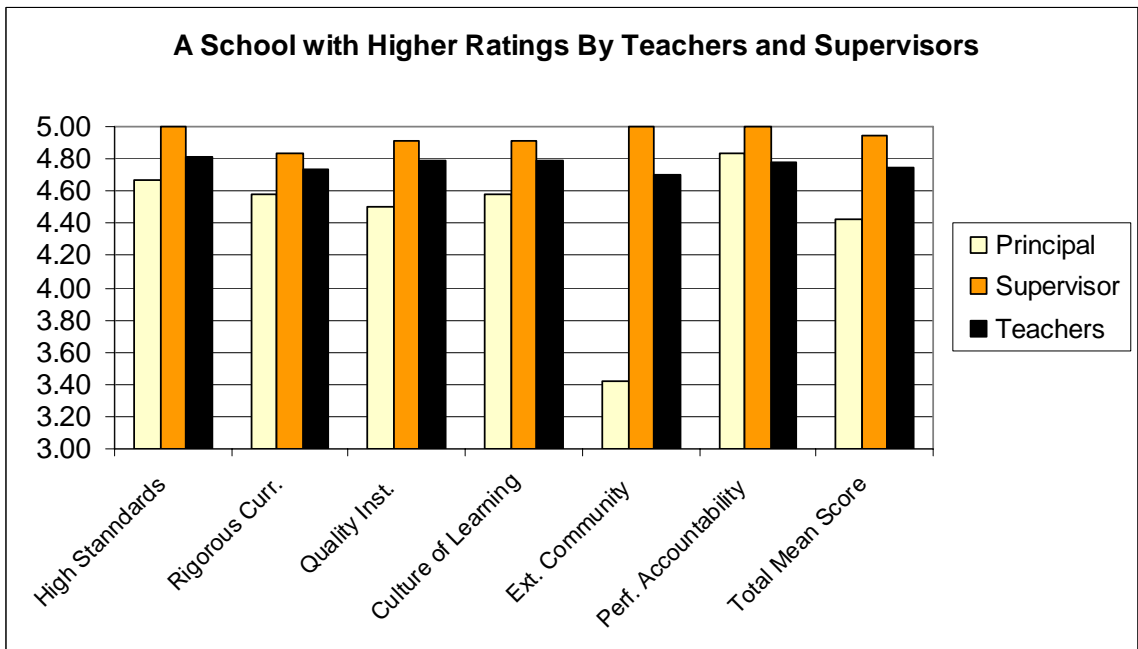Figure 15: School Rating Comparison 1

Figure 16: School Rating Comparison 2

The between-school variation of the scores shows 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 point to possible associates between school characteristics and leadership assessment results. From the comparison provided between School 1 and School 11, one may hypothesize that high performing schools tend to have not only higher average ratings but also higher inter-rater correlations among principal, the supervisor and the teacher ratings for leadership effectiveness. On the other hand, other factors may be at play, school type, principal experience, and 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. Just as we see here, a low inter-rater correlation of the overall ratings among principals, supervisors and teachers does not necessary diminish the reliability of the instrument until we look at how the rating consistency vary in different settings.

Performance Nomination Summary

As a step toward measuring the criterion validity of the VAL-ED instrument as an assessment tool for principal performance, ratings for the four-dimension standards that are used widely by the Chinese school systems for principal evaluations were compared with the VAL-ED results. Recall that the four dimensions, value, ability, diligence, and achievement are broad categories that are used to evaluate educational administrators

136

such as school principals. The four dimensions are measured by four items on a 1-5

Likert scale as an addition to the VAL-ED assessment in this study. These four items

form the scale for the "Chinese standards." The mean score of the four items is used for

the comparison with the mean score of the 72-item VAL-ED assessment.

The correlations between the Chinese standards mean scores and the VAL-ED

mean scores were obtained at the school level and presented in Table 20. The unit of

analysis is the school average score by the teachers, the mean score of the supervisor, and

the principal self-assessment for each school (N=18). Correlations are high between the

VAL-ED scores and the scores for the Chinese standards within each rating group: 0.84

by the teachers, 0.68 by the principals, and 0.79 by the supervisors. The cross-group

correlations between the VAL-ED and the Chinese Standard scores are much lower.

With the exception of 0.39 between the VAL-ED principal scores and the Chinese

standard scores by the supervisors, the others are all below 0.30.

Table 20: VAL-ED and Chinese Standard School-Level Mean Score Correlations (N=18)

	VAL-ED Ratings			Chinese Standard Ratings		
	Teacher	Principal	Supervisor	Teacher	Principal	Supervisor
VAL-ED Teacher	1.00					
VAL-ED Principal	0.00	1.00				
VAL-ED Supervisor	0.04	0.15	1.00			
Chinese Teacher	0.84	0.11	0.25	1.00		
Chinese Principal	0.00	0.68	0.28	0.19	1.00	
Chinese Supervisor	0.16	0.39	0.79	0.34	0.43	1.00

The correlations provide the initial evidence that when rated by the same group of

raters, i.e., teachers or principals, the ratings on the principals exhibit moderate to high

correlations between the VAL-ED ratings and the ratings using a different set of criteria,

the Chinese four-dimension standards measuring values, ability, diligence, and achievement.

The relations of the two sets of criteria were further examined by plotting the quartile distribution of the VAL-ED mean scores by the rating category of the Chinese standards (Figure 17, 18 and 19). Recall that the four rating criteria, value, ability, diligence, and achievement, were listed as four assessment items and rated on a Likert scale of 1-5 as an additional component of the principal assessment. To graph 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, >=2 and <3 for Level 2, >=3 and <4 for Level 3, >=4 and <=5 for Level 4), individual VAL-ED mean scores were sorted by these levels and plotted within each level.
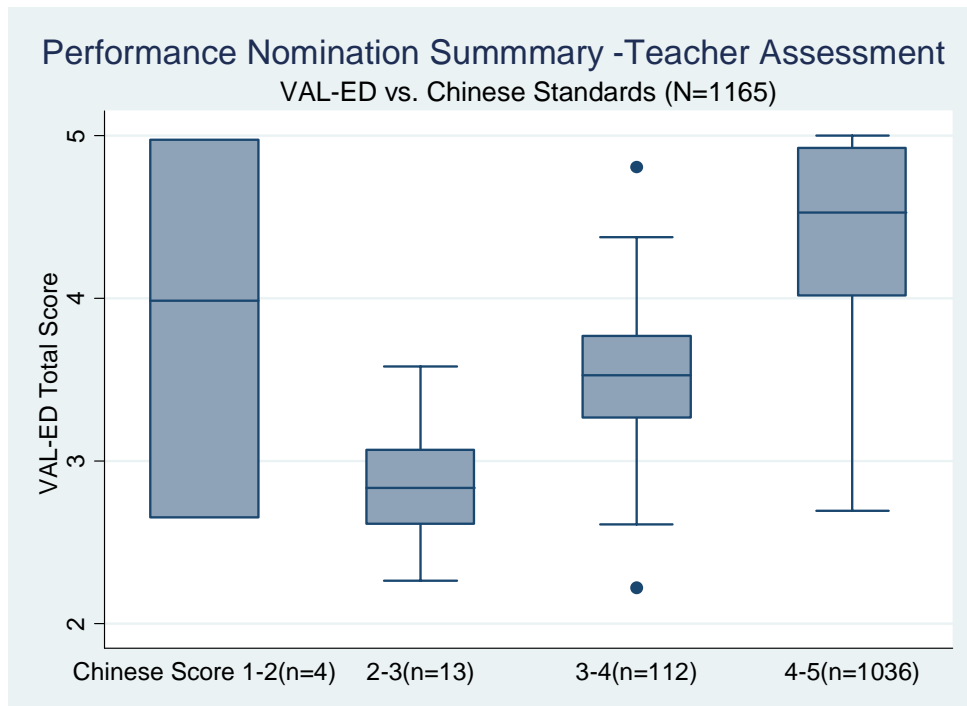

Figure 17: Criterion Validity Check Using the Chinese Standards –Teacher Assessment
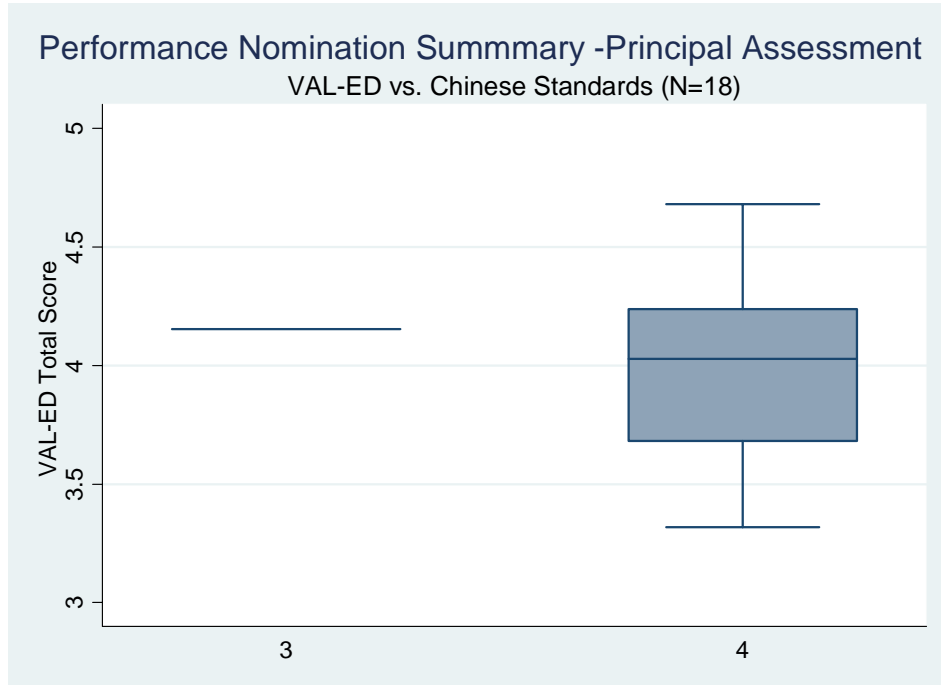
Figure 18: Criterion Validity Check Using the Chinese Standards –Principal Assessment
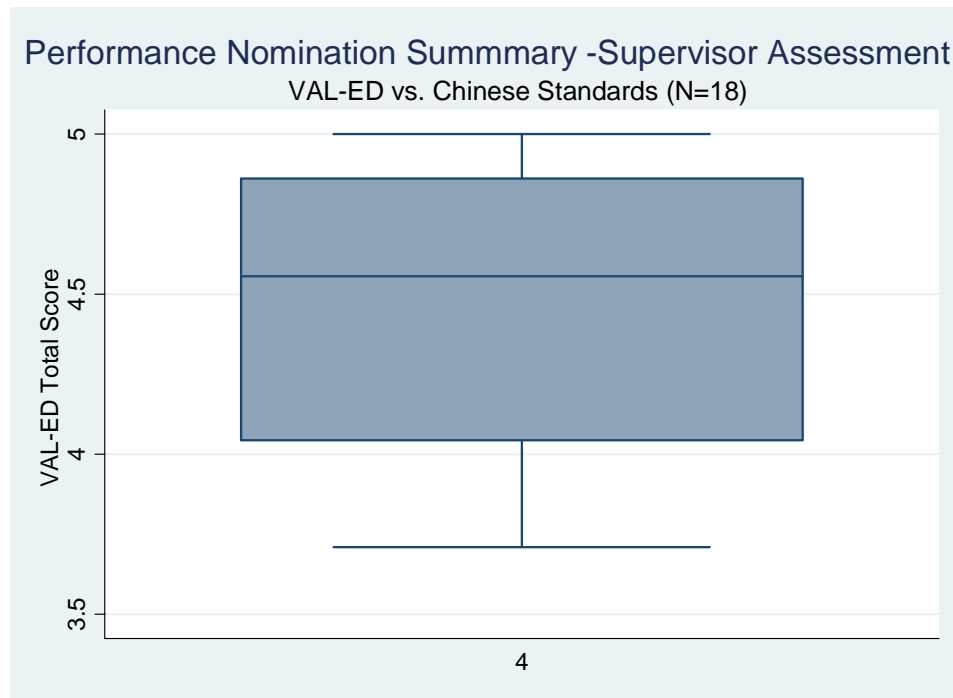

Figure 19: Criterion Validity Check Using the Chinese Standards
–Supervisor Assessment

The results are more interesting with the teacher assessment scores with the large number of observations (N=1165) and demonstrated variation (Figure 17). The first box plot has only four observations with average Chinese standard scores below 2.0 and the VAL-ED scores are spread out between 2.6 and 5. As discussed previously, the ratings were highly skewed to the right with most of the scores above average, and very few ratings fell below 3.0. The other three box plots for the VAL-ED scores are nicely aligned in an upward pattern, reflecting the increase of the Chinese standard scores. The largest group belongs to Category 4 where the Chinese standard scores are between 4 and 5 with 1036 observations. In this group, observations between the 75th percentile (upper hinge) and the 25th percentile (lower hinge) take practically the full range between 4 and 5 on the VAL-ED score line, and the median line is at 4.52, almost a perfect mid point between 4 and 5. Observations below the 25th percentile, however, have VAL-ED scores ranging from 4.01 to 2.69, indicating that the VAL-ED scores cover a wide range of variation in teachers' perception of their principals' effectiveness. The same can be said about the two box plots for Category 2 and Category 3, while the Chinese standard ratings match the box positions of VAL-ED scores representing observations above the 25th percentile and below the 75th percentile, the whiskers of the boxes illustrate wider ranges of the variation in the VAL-ED scores than those for the Chinese standards.

Because there are only 18 observations for the principal and for the supervisor assessments, information from the box plots (Figure 18 for the principal assessment and Figure 19 for the supervisor assessment) are thus limited. Only one principal has a self average rating in Category 3 and the rest of 17 are in Category 4. Among the 17, the median line is at 4.02, as compared with 4.52 median by the teachers in the same

category. The range of the box is from 3.68 to 4.23 for the VAL-ED scores. In other words, principals who rated themselves between 4 and 5 using the Chinese standards rated themselves from 3.68 to 4.23, in a comparatively lower range. As for the supervisor ratings, there is little discrepancy between how the supervisors rated the principals using either set of standards. The Chinese standard ratings are all in Category 4 and the box is well situated between 4 and 5 for the VAL-ED scores. However, the whisker goes down to 3.71, showing again a wider range of the VAL-ED scores.

In psychometrics, *criterion validity* is a measure of how well one variable or set of variables predicts an outcome based on information from other variables. Criterion validity of an assessment can be depicted by concurrent or predictive validity. In this case, the concurrent validity of the VAL-ED items was gauged by comparing their results to the Chinese criteria, assessed simultaneously. The results show that principals that received high scores based on the VAL-ED scale tend 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 provides evidence for the criterion-validity of the VAL-ED as an assessment tool for principal effectiveness.

Evidence Used for Effectiveness Rating

The sources of evidence used by the respondents were examined as an additional check for the reliability of the rating scale results. The percentage of observations that did not provide any source of evidence for a rating scale was calculated for each item. For the teacher assessments (Table 21), the missing-evidence percentages of the 72 items range from 7.72% to 12.27%, with a standard deviation of .01, and the average is 10.33%.

Core component 4, Culture of Learning and Professional Behaviors, has the lowest

missing average at 8.54%; Core component 5, Connections to External Communities, has

the highest missing percentage at 11.30%. The missing data rate is tolerable and no items

jump out as having particularly high percentages of observations that missed evidences.

For the supervisor assessment (n=18) and the principal self-assessment (n=18), the

percentages of observations that missed evidence are much lower. The average missing

evidence percentage is 5.56% for the supervisor assessments and 4.15% for the principal

assessments.

Table 21: Percentages of Observations Missing Sources of Evidence
(Teacher Assessment, N=1165)

Component 1		Component 2		Component 3		Component 4		Component 5		Component 6	
Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing	Item No.	Pct Missing
1	10.13%	13	10.39%	25	10.27%	37	7.90%	49	10.73%	61	10.13%
2	10.13%	14	10.64%	26	10.52%	38	7.73%	50	10.73%	62	10.73%
3	10.64%	15	10.56%	27	10.44%	39	7.73%	51	11.16%	63	10.47%
4	10.73%	16	10.73%	28	10.61%	40	10.99%	52	10.64%	64	10.64%
5	11.42%	17	10.21%	29	10.09%	41	8.67%	53	11.50%	65	10.73%
6	10.82%	18	10.21%	30	10.09%	42	8.07%	54	11.24%	66	10.56%
7	10.56%	19	10.64%	31	10.52%	43	8.33%	55	11.76%	67	10.56%
8	10.99%	20	10.13%	32	10.01%	44	9.96%	56	11.93%	68	10.30%
9	11.07%	21	10.39%	33	10.27%	45	8.41%	57	11.33%	69	10.99%
10	11.59%	22	10.39%	34	10.27%	46	8.15%	58	11.07%	70	10.73%
11	11.33%	23	10.21%	35	10.09%	47	8.58%	59	11.24%	71	10.56%
12	10.64%	24	10.47%	36	10.35%	48	7.98%	60	12.27%	72	10.90%
Mean	10.84%		10.41%		10.29%		8.54%		11.30%		10.61%

Results on the types of evidence used for effective rating (Table 22) show that the

teachers were most likely to choose "personal observation" as the evidence (74% of the

items), with a distant second choice of "school document" at 28%. Compared with the

teachers, the principals and supervisors were more likely to use documents than observations. More than 60% of the principal self-ratings and supervisor ratings used "school documents" as evidence, and "report by others" was more frequently used by the supervisors than by the principals and teachers, at 42%.

Table 22: Types of Evidence Used for Effectiveness Rating

	Teacher	Supervisor	Principal
Report from others	25%	42%	35%
Personal Observation	74%	52%	54%
School documents	28%	62%	65%
School projects or activities	24%	33%	41%
Other Sources	5%	6%	15%
No Evidence	1%	2%	0%
Average # of sources	2.12	2.50	3.10

Overall, about 90% of the effectiveness ratings were supported by one or more evidences and the pattern of the pattern of the types of evidences seem to be reasonable and consistent. Including sources of evidence is an important feature of the VAL-ED assessment aiming at adding accuracy and reliability of the results. The findings from examining the evidences used for the effectiveness ratings in the 19 Chinese schools demonstrate that the users responded relatively well with the requirement of checking on the sources of their decisions.

CHAPTER V

DISCUSSION

In this chapter, the research findings are discussed addressing each research question, including both the substantive and methodological aspects of testing a leadership theoretical framework and the assessment instrument cross-culturally. Upon discussing the limitations of the studies, specific suggestions on possible next steps to further the research agenda are presented. The discussion chapter ends with thoughts on the significance and implications of the research findings for cross-cultural studies in the field of educational leadership.

Interpretation of the Findings

This dissertation sets out to answer three research questions. First, how well does the Learning-Centered Leadership Framework (Porter et al., 2006), 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, is there evidence that the instrument has construct validity, and yields consistent results when taken by the targeted participants? Third, based on the results of the first two questions, should the framework and the instrument be modified, and if yes, what may be some suggestions to enhance the cross-cultural relevance and utility of the Learning-Centered Leadership Framework and the VAL-ED for Chinese principals? The empirical studies were designed to address the first two research questions, which examine the cross-

cultural fit of the leadership model from two angles: the alignment of the theoretical

framework with practices and standards in China, and the construct validity and

reliability of the assessment instrument in urban Chinese schools.   The third research

question, assessing the relevance and the utility of the framework and assessment

instrument, can only be answered based on the analytic findings of the first two questions.

Research Question One: The Alignment of the LCL Framework with Chinese Standards
and Practices

The alignment ratings on the importance and relevance of the VAL-ED leadership

behavior indicators given by the expert panel illuminated the commonalities shared by

the U.S. and Chinese educational systems and the differences between the two systems in

identifying and practicing important leadership behaviors that enhance learning.

Findings from the expert-panel alignment study provided evidence that the

Learning-Centered Leadership Framework is well aligned with what are considered

important for the success of schools and students to improve learning. The expert panel

agreed that the intersection of the two dimensions (core components and key processes)

covered the domain of school leadership that might influence school effectiveness,

despite concerns over the appearance of redundancy in some items related to monitoring

and communicating, and the lack of mentioning of motivation and focus on teacher

professional development.

However, the alignment resided more with the perceived standards than with the

current practices of school principals in the opinion of school administrators and

researchers in educational leadership development. When ratings were aggregated and

examined as group averages, experts on the panel consistently rated the importance of the six components at a level that was higher than neutral, but on a level that was barely above or below neutral when it came to the rating for current practices. Such difference between what was being done now and what it should be was perceived consistently by both researchers and practitioners on the expert panel, only that the "reality gap" was slightly bigger in the eyes of practitioners than the researchers.

Among the findings, the existence of the "reality gap" is most informative. The discovery of what is lacking may provide the necessary push for establishing the knowledgebase and the systemic support for principal professionalization (Chu, 2003). Future research may focus on developing a deeper understanding of the sociopolitical nature of the gap, and more importantly, on the establishment of professional standards and training programs that will close the gaps.

As a first step to better understand the nature of the alignment and the "reality gap", this study did not stop at the broad statement on the theoretical alignment of the framework perceived either by the current practices or the perceived importance, but explored ways to examine the alignment in a multifaceted approach: first at the level of the framework elements (core components or key processes), and then at the item level. The variations among the core components revealed that principals leadership behaviors for Rigorous Curriculum, Quality Instruction, and Culture of Learning and Professional Behaviors appeared to have smaller gaps between what are considered important and how much are in practice. In comparison, what the principals were doing to connect with the external communities seemed to lag behind the level of the perceived importance of this core component. At the item level, items that had the largest gaps between relevance and

146

importance illustrated that there were noteworthy deficiencies in leadership behaviors for advocating and communicating, which are the key processes emphasizing serving the diverse needs of students and connecting with both the internal and the external school community. Such findings on the variation of alignment between the VAL-ED core processes and key processes with the standards and practices of Chinese school leadership prompted comments, questions and suggestions from the expert panel that emphasized the need of further inquiries into the social, cultural and policy factors that impact the professional standards and performance accountability measures for school principals, which would undoubtedly be the driving forces for leadership behaviors in Chinese schools.

Research Question Two: Instrument Construct Validity and Reliability

The combined efforts of the cognitive interviews and the large-scale assessment in 19 schools provided convincing evidence that the instrument has strong internal consistency, that is, it yields consistent results when taken by the intended participants. There were also sufficient evidences to confirm that the translated version of the VAL-ED, with its content and format largely intact, has content validity in that the items largely reflect important aspects of the leadership behaviors of principals. In addition, the VAL-ED exhibited important criterion validity as an instrument for leadership assessment when compared with the Chinese standards. However, the studies also pointed out important areas of concerns for construct validity. The content validity and discriminatory trait of the exiting items, at least based on the results from the 19 schools, do not completely support the conceptual framework.

Confirming the basic alignment of the Learning-Centered Leadership Framework with how school leadership is perceived in China was an important first step in establishing construct validity of the assessment instrument. The construct validity of the VAL-ED assessment items was further examined in the two studies that followed. The two studies focused on the content validity and criterion validity of the instrument.

To be specific, the cognitive interviews of two school directors, two principals, and two teachers provided opportunities for in-depth inquires into how items are interpreted by the respondents. Findings from the cognitive interviews indicate that a majority of the VAL-ED items in their original form could be easily understood by the respondents, and the intent of the items were clear and agreed upon. However, the respondents pointed out that not all of the items fit the Chinese educational context. The misfit might appear in two forms: (a) the leadership behavior cannot occur in the local context due to differences in policy or other cultural reasons thus might be irrelevant; (b) the leadership behavior is in practice and important in the Chinese context, but the intent of the item is interpreted differently, in most cases reflecting the differences in the directions of educational reform initiatives of the United States and China. For example, while leadership behaviors for Rigorous Curriculum were considered important and well reflected in the action of Chinese principals, some items may be missing the point of the new reform initiative by the Chinese Ministry of Education on "quality-oriented education" that puts its emphasis on adjusting the rigor of academic content to make room for character-building, physical health, and other non-academic developmental priorities. The ratings for these items may turn out to be high but will not have good construct validity in that the do not cover the true intent of the measurement.

In addition to providing important insight on the content validity of the VAL-ED items, the cognitive interview respondents gave overall confirmation of the assessment administering process, instrument structure (instruction, effectiveness rating scale, and sources of evidence) with suggestions for small adjustments.

The construct validity of the instrument was examined first by accounting for the common variances of the 72 items through exploratory factor analysis, and then by measuring the criterion validity using a set of Chinese standards for principal performance. The purpose of the EFA analysis is to find out how the common variances of the items cluster based on the real assessment scores and if such clustering reflects the scales of core components and key processes that are set up based on the theoretical framework. Results from the EFA show that the first four core components were more distinctively represented by the identified factors than the last two. And the key process scales were less identified by the factor structure than the core components. There is a noticeable concentration of common variances onto a factor that is comprised of most of the items for the last two components (Connections to External Community and Systemic Performance Accountability) and a significant portion of the item for the last two processes (communicating and monitoring), an indication that the constructs represented by the subscales lack discriminate validity. Interestingly, the attention to these areas coincides quite well with the concerns over the appearance of redundancy brought up by the expert panel members and the cognitive interview respondents. Recall that the missing data pattern (Table 10 and Table 11) also show that missing data rates are relatively higher in items for the last two components and the last two processes. Although the exact reason why higher missing data rates occurred in these subscales

cannot be pinpointed within the scope of this study, the occurrence adds to the findings of the qualitative feedback and the EFA results, which call for attention to the elements of the assessment that appear to be problematic to the Chinese respondents.

A possible explanation to the clustering of items of the last two components and the last two key processes is "bleeding" between the two dimensions. One may argue that connecting to external communities is more of a process than a component. It can be regarded as one of the venues of communicating for the purpose of enhancing core learning-centered components such as standards, curriculum, instructions and learning culture. One may also argue that the component of Systemic Performance Accountability shares certain common connotation with the process of monitoring, where accountability serves as a mechanism utilized during monitoring. Instead of a learning-centered component, performance accountability may be a tool used in the processes to ensure that curriculum, instruction, and professional behaviors are taking place.

Overall, the EFA results provided empirical evidence that the VAL-ED two-dimensional and 32-cell structure can only be confirmed partially. The evidence revealed areas in the framework that the current items, upon translation, did not have the discriminatory power to identify and set apart.

The criterion validity evidence for the instrument score was measured by comparing the VAL-ED scores with the scores using the Chinese criteria for leadership comprised of four dimensions: value, ability, diligence, and achievement. The results were very encouraging. When comparing the two sets of criteria within each rater group, the VAL-ED scores have high correlations with the scores of the Chinese standards, and

150

the distributions of two sets of scores are well aligned to reflect the continuum of rating scales.

Study 3 also amounts quantitative evidence on the reliability of the VAL-ED assessment based on the assessment scores from principals, supervisors and teachers. The VAL-ED demonstrated excellent internal consistency at the scale level for the core components and key processes, and as a 72-item instrument overall after being translated into Chinese and administered in urban Chinese schools. Several other aspects of the instrument reliability were also examined and the results reveal more complex findings. First, the internal consistencies among the components and among the processes appear to be the highest with the teachers and the lowest with the principals. In other words, when doing self-ratings, principals are more likely to treat the items in different components differently and the scores will yield lower correlations than the teachers and the supervisors. Second, the inter-rater correlations among the three parties involved in the assessment, principal, supervisor and teacher, are very low, meaning that the "360-degree" assessment process yielded different results from the participants. It was found that the inter-rater correlations appeared to be higher with higher performing schools. But considering the limited sample size (N=18), this is a very preliminary finding that calls for further inquires.

In summary, the VAL-ED assessment instrument was proven to be a tool of strong internal reliability, and the cross-cultural validity of the VAL-ED could be partially confirmed through the examination of content validity and the criterion validity evidence. However, the discrepancies between the VAL-ED framework and the results obtained from the qualitative interviews and quantitative empirical data were also

noteworthy. Understanding such discrepancies will be the critical gateway into determining whether the VAL-ED should be and can be improved upon to fit in a different cultural setting.

Research Question Three: Relevance and Utility of the Framework and the Instrument

Findings from the three empirical studies gave support to the claim that the Learning-Centered Leadership Framework is very relevant to Chinese urban schools. Input from the Chinese researchers, principals, school directors and teachers consistently confirmed that there is a strong cross-cultural 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.

For the VAL-ED to be a useful tool as an assessment instrument of principal effectiveness for student learning, however, the theoretical framework and the assessment items will need to be modified to reflect the differences between the two educational systems that are rooted in a wide range of factors such as economic and demographic condition, governance structure, and how student achievement has been defined and measured in the recent decades. While the two-dimension conception of leadership was well received by the Chinese participants of the studies, some aspects of the core components and key processes might become clearer conceptually and the instrument items more robust if more local contextual factors are taken into consideration.

The finding of no complete cross-cultural fit of the VAL-ED theoretical framework and assessment without further modification is expected. Well grounded in

school leadership literature, the validity and reliability of the VAL-ED assessment are still going through large-scale field tests so that more evidence can be obtained on the predictive power of the instrument for effective principals and successful schools. Even with the assumption that the current VAL-ED framework and instrument are perfectly suitable for U.S. schools, the major educational reform efforts taking place in China that emphasize social learning, creative thinking, and physical health need to be covered as the priorities of the Chinese basic education in the content domain of the assessment for it to be meaningful and useful. Adding to the complexity, 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 & Cravens, 2008). Performance goals established based on professional standards that do not have systemic support will be impossible for principals to reach, 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. Ideally, the adapted theoretical framework for the Chinese setting will align with the U.S. version on the level of broad definitions of leadership and reflect the common goals of education, and it is also equipped to capture deeper differences in the "directional focus" of the leadership behaviors that may lead Chinese principals to improved student academic and social learning.

Suggestions for Future Research

Findings of this paper provided opportunities for future research to further explore the possibility of adapting the learning-centered leadership framework and the assessment instrument with cross-cultural validity and reliability theoretically and empirically. Preliminary suggestions for developing a learning-centered leadership assessment system that fits the needs of Chinese urban schools are made in three parts. First, the limitations to this paper are pointed out as areas that can be improved upon for future studies; second, steps of modifying the VAL-ED framework, and testing instrument validity and reliability in the Chinese setting are discussed; third, possible research inquiries using the new instrument to inform educational reform, enhance principal professional performance, and measure school improvement are proposed.

Limitations to be Addressed in Future Studies

The extent and the depth of the studies 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.

First, there are some regrettable flaws to the empirical studies. For the expert panel study, the proportion of practitioners should have been higher than four out of 12, knowing now that the input from principals and school directors were well reflected through out the rest of the studies. For the cognitive interviews, ideally there should be at least two iterations so that the second group may provide further reflection regarding the modifications of instrument content and structure based on the feedback of the first group. Instead, the modifications to the instrument did not receive further review from the

respondents before the assessment effort had to be started in the schools. For the large-scale assessment study, there is a general concern over the lack of attention given to understanding the definition of the core components and the key processes in the framework, and to the instruction of the assessment by the participants. It was observed that, although offered with sufficient time, some teachers appeared to spend very little time to read the instruction of the assessment and some finish the assessments in less than 20 minutes. The results of the assessment could be seriously compromised when a teacher did not read the instruction well and rush to make selections for effectiveness rating and sources of evidence. It was explained by some of the local researchers that schools nowadays have to respond to multiple surveys from various entities and the teachers may suffer from survey burnouts. Also related to Study 3, the generalizability of the results from the 19-school sample is limited. Guangzhou is one of the largest cities in China and has been on the forefront of economic development in the recent decades. Its local educational management structure and policies could be quite different from other smaller cities despite the national curriculum standards and centralized governance scheme. Even within the city, a random sampling process could have yielded a school sample conceivably more representative of various types of principals and schools. Because the principals volunteered to participate in the studies, selection bias is possible. Despite the variation in the final sample in terms of important school indicators such as school size, type, and percent of senior teachers, the principals in the study may have represented the upper bound of effective leadership behavior range. This possible upward bias may have furthered the skew of the VAL-ED scores.

On the theoretical front, the difficulty of measuring leadership using a translated instrument that had to keep its original framework and content manifested in several ways. Measuring the effectiveness of a principal in itself is sensitive subject and an act that is carried out in a wide range of forms by the Chinese schools (Zhao & Wang, 2007). The quality of the assessment study required the teachers to understand (a) the purpose of the research; (b) that the results won't be used as the real administrative review of the principal; (c) that the results are completely anonymous; and (d) that the results will be meaningless to the research if the teacher does not fill it out as if it were a real event of assessing the principal. However, the extent to which these cognitive requirements on the teachers were met was very hard to gauge.

Even under the assumption that the participants understood the task and cooperated fully, the studies had to face the challenge of finding the balance between the technical need of maintaining strict construct equivalence to measure the cross-cultural fit and the practical necessity of having an instrument that has items that are applicable in the local context to reduce the rate of null responses. At the onset of the translation process, small changes to the meaning of several times were made so that the items did not become completely irrelevant in the Chinese context which could in turn unravel the 72-item structure. Subtle changes were made to address very specific differences in the two systems, e.g., translation for "special needs" was broadened to include students that lag behind in learning, while the term is often used specifically for students with disabilities in the U.S. educational setting. However, these changes are not the real threat to validity because they were kept at minimum and were only done if the original intent of the items was not altered. The real concerns involved items that were identified by the

expert panel and cognitive interview respondents as ambiguous or problematic to be used to measure principal leadership behaviors. These items were kept in the assessment instrument to keep the original instrument intact. However, it is a concern that the assessment scores can only provide a partial picture of the cross-cultural construct validity of these items. For example, items for the core component of Rigorous Curriculum, which may receive consistent high ratings because the focus on academic learning by the items is well reflected by the current actions taken by principals and teachers in the schools. But the construct validity of these items is questionable because they do not reflect the intent the new educational reform that calls for curriculum design and implementation that are more locally driven with less emphasis on achievement scores more on "soft skills" such as teamwork and creativity. On the other hand, there are other items that are considered important by the practitioners and are being promoted by new policies, but are still not the common practice in urban Chinese schools due to reasons that are external and environmental, received comparatively low ratings or had missing data problems. Naturally, the core conditions for learning-centered leadership such as principal leadership and teacher professional community could vary to a great extent by school. And schools in the sample may very well be at different stages of adapting to new policies. The impact of contextual complexity on the construct validity of the VAL-ED is undoubtedly immense when considering the magnitude of the Chinese basic educational system.

Framework: A modified framework is proposed (Figure 19) for the following

purposes: (a) to accentuate the first four core components that are learning-centered, (b)

to address of "bleeding" issue between the last two core components and key processes;

and (c) to stay consistent with the two-dimension structure of the U.S. framework.

Learning-Centered Components	Leadership Processes					
	Planning *With Shared Vision and Objectives*	Implementing *Actions for Student Performance*	Supporting *Teachers for Continuous Improvement*	Advocating *For All Students*	Communicating *With Internal and External Communities*	Monitoring *Systemic Performance Accountability*
High Standards for Student Performance						
Rigorous Curriculum (content)						
Quality Instruction (pedagogy)						
Culture of Learning & Professional Behavior						

Figure 20: Learning-Centered Leadership Framework for Chinese School Leaders Adapted from the Learning-Centered Leadership Framework by Porter et al., (2006)

The adapted framework keeps the first core components: High Standards for

Student Learning, Rigorous Curriculum, High Quality Instruction, and Culture of

Learning and Professional Behaviors. These four components distinctively address the

areas of student academic and social learning that must be the center of school leadership.

The last two core components from the original framework are now incorporated into the

key processes of communicating and monitoring. Feedback and evidence from the studies of this paper show that connections to external communities and systemic performance accountability are more process-related than as stand-alone components for student learning. Connecting to parents and other entities is a form of communication and an action that should be taken regarding any and all of the first four core components. The same logic applies to systemic performance accountability, where it is more of a monitory mechanism ensuring that the goals for the first four core components are achieved.

The importance of communicating with parents and other important external entities cannot be overlooked, however, as a leadership behavior considered important but hardly done by the Chinese practitioners. To make sure the key processes are clearly defined, each key process is now presented in the framework with a modifier for the action (also see Figure 20). The modifiers are based on the original definitions of the key processes but have incorporated the findings from the studies. Specifically, supporting is qualified as "supporting teachers for continuous improvement" to cover the suggestion made by the expert panel on teacher professional development and motivation for improvement; communicating is qualified by "with internal and external communities", and monitoring is qualified by "system performance accountability", integrating the last two core components of the original framework. Just as in the original framework, each core component will intersect with the six key processes. For example, Rigorous Curriculum as a learning-centered core component of principal leadership is measured by how the principal ensures that the six processes are enacted, including communicating

with internal and external communities about rigorous curriculum, and monitoring systemic performance accountability for rigorous curriculum.

Rating Scale: With a 4x6 cell structure, the instrument can have three items for each cell and still have a total of 72 items. The increased number per cell gives the cells increased domain coverage for each intersection of a core component and a key process. The original items will be examined and modified based on the feedback and suggestions made by the expert panel and interview respondents to better reflect the Chinese policy and social context, including important aspects such as the focus on student social learning in curriculum planning and instructional practices, providing motivation for teacher professional development, and working with parents and supportive organizations outside of the schools.

All elements of the adapted framework and assessment instrument will be put through the validity and reliability examination process.

Empirical Study Design: The study design of this paper was modeled after the U.S. pilot and field studies for the VAL-ED (Porter et al., 2008) with some small adjustments. For example, the pilot study for VAL-ED included a sorting study where a small number of principals were asked to put items back into the cells of the two-dimensional framework. This exercise was not done for this paper considering that with limited time, resources and a small sample of principals available for several hours away from their work schedules. It was also decided that more in-depth information could be obtained from the cognitive interviews on the content validity and instrument utility than from the sorting study.

Overall, the three studies worked well to collect qualitative and quantitative evidence for the examination of validity and reliability of the framework and the assessment instrument. Future studies for the adapted framework and the assessment instrument may benefit from the following adjustments while still follow the basic design for the empirical studies.

First, the expert-panel will be given more time to study the background information on the rationale for the newly adapted framework and its link to research including the findings of this paper. The ration of practitioners on the panel can be increased to at least 50% of the sample. The alignment form was proven to be very informative but it the results may be more enlightening if there could be face-to-face discussions with the panel members about their feedback and suggestions.

After incorporating the input of the expert-panel, the instrument can go into the second phase for cognitive interviews. At two rounds of cognitive interviews with two different groups of principals, school directors and teachers should be conducted to add to the iterations of improving the construct validity of the items. A sorting study with a third group of participants can be added after the instrument has gone through the cognitive interviews and further modified, before it goes to the schools. The results of the sorting study at this point will be meaningful in that the conception of where the items belong can be compared with the initial factor analysis results from the previous assessment scores and to see if the items can be more distinctively identified to fit the framework.

The assessment study for instrument validity and reliability can be improved by using a randomize selection process for schools that are more representative of urban

Chinese schools, in multiple cities of different sizes and economic conditions. It may

also benefit from a two-step process where a small portion of the sampled schools will be

assessed first. The results can be analyzed with exploratory factor analysis to see if the

factor loadings better reflect the new framework structure, and if further modifications to

the items are necessary. The second round of assessment effort will then used the update

version and its results can be analyzed with confirmatory factor analysis.

New Research Inquiries

It is foreseeable that five to 10 years from now, with the establishment of sound

professional standards and active principal training based on the standards, the "reality

gap" will become smaller as the effectiveness of learning-centered leadership gets

stronger. Such assumption might serve as an excellent premise for new research inquires

that explore the causal relationship between enhanced learning-centered leadership

behaviors and the improvement of student learning over time. Two research questions

may be proposed:

Research Question 1 – Does professional development for learning-centered

leadership lead to higher effectiveness ratings on the principal assessment (the modified

VAL-ED)?

Research Question 2 – Does improvement in learning-centered leadership

behaviors measured by the modified VAL-ED lead to improved school performance

measured by student social and learning indicators?

Several important conditions have to be met before a research study investigating

the impact of learning-centered leadership on student social and academic learning can

take place. The satisfaction of these conditions requires an extensive web of alignment among the following key elements:

1. Professional standards for school principals that are aligned with the new priorities of "quality-oriented" educational reform;

2. Definitions of effective schools with core performance measures that are also aligned with the new priorities of "quality-oriented" educational reform;

3. Development and validation of a leadership assessment instrument that is aligned with the professional standards;

4. Development and evaluation of principal training programs that are aligned with the professional standards and helping principals gain competencies that can be measured by the leadership assessment instrument.

Figure 21 illustrates a preliminary research design addressing the research questions with the assumption that the above mentioned conditions are met. Without delving into a full-blown description of a new study, the design can be summarized by the following key features:

1. A randomized field experiment design will be employed, where two groups of randomly selected principals with statistically equivalent characteristics will be assigned an experiment group and a control group. Principals in the experimental group will receive training that is specifically designed to be aligned to the new professional standards.

2. Both groups will be assessed by the modified VAL-ED twice a year.

3. The schools where the principals work will be assessed twice a year using performance measures that include core student social learning and academic

learning indicators. The assessments on student learning will start one year

after the principal training occurs to ensure temporal order of cause and effect.

4.   To measure longitudinal effect on change, the experiment will be for at least
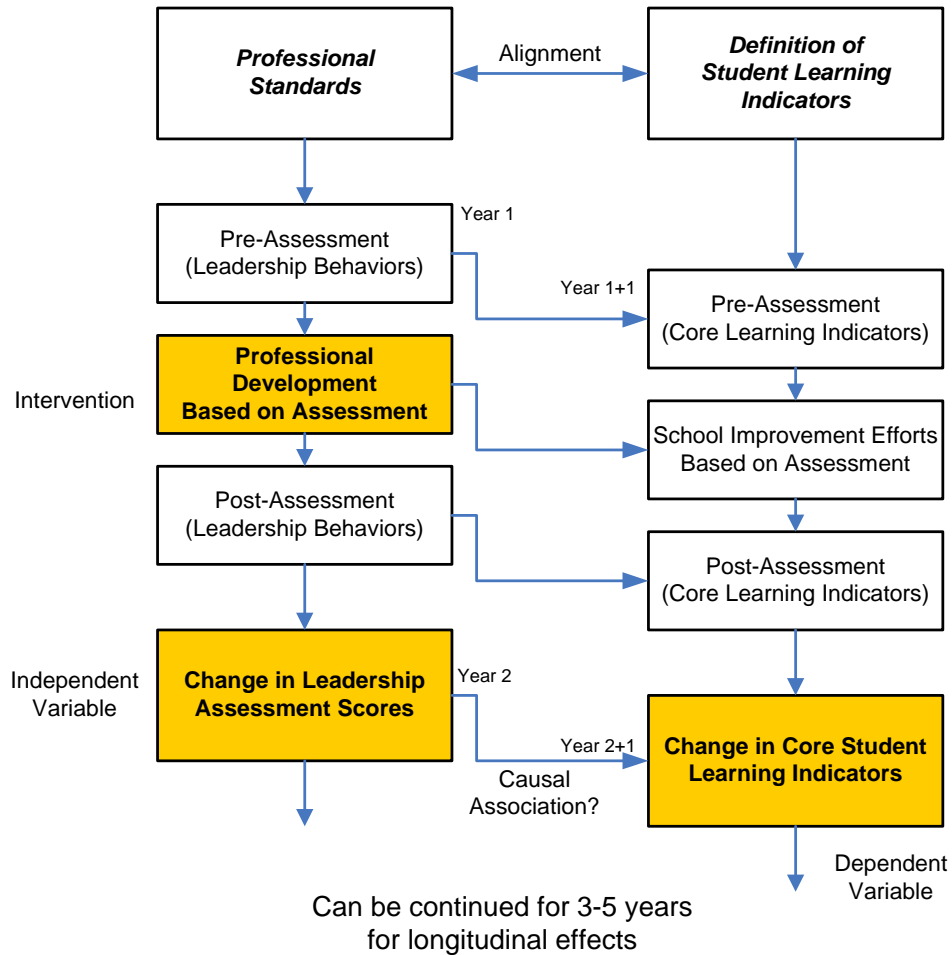
three years.

Figure 21: Investigating the Impact of Learning-Centered Leadership
on Student Social and Academic Learning

The results can be analyzed to test two hypotheses. Hypothesis 1 states that the

increase of the VAL-ED score for the experimental group that receive training is bigger

than that of the control group, and may be tested with methods such as ANOVA to see if

ΔVAL-ED$_{(experimental)}$ > ΔVAL-ED$_{(control).}$ Hypothesis 2 states that change in the VAL-ED

effectiveness rating has a positive impact on student social and academic learning results.

Methods such as ordinary least squares (OLS), hierarchical linear modeling (HLM), and

structural equation modeling (SEM) all have their strengths and weaknesses in

identifying associations and causal effects when dealing with the complexity of

educational outcomes. Which statistical methods should be employed may very well

depend upon how the student learning variables are defined and collected.

The proposed research questions are only examples to demonstrate the immense

need and possibilities for future scholarly work in the field. It cannot be emphasized

enough that much more foundation-building has to be done before an assessment

instrument for leadership can be truly useful in promoting student learning.

Implications

Today schools in the U.S. are mandated by the government and the public to

operate with higher academic standards and more performance accountability. Facing the

threat of dwindling manufacturing jobs, trade deficit, and the economic growth of nations

such as China, the No China Left Behind legislation of 2001 anchors its efforts for

improving the quality of education on raising student academic achievement, measured

by standardized testing results in core subjects such as mathematics, reading and science.

While the 90,000 or so school districts in the U.S. still maintain the local control over

important decisions such as instructional personnel, classroom size, and extracurricular

activities, on accountability issues, there is an over shift toward great standardization of

curriculum, testing, teacher training, and performance monitoring (Guthrie & Wong, 2007).

Meanwhile, schools in China are criticized for failing to prepare students with critical thinking skills and creativity for a competitive global economy. Educators are raising serious concerns over the consequence of the tightly-controlled, test-based education system that only prepares students to be proficient in exams but does not enhances creativity, problem-solving, and other important qualities. The Chinese curriculum reform, also started in 2001, promotes new curriculum content and instructional practices that are localized and focus on the intrinsic value of knowledge. The reform also mandates the schools to have sufficient hours of physical activities, to reduce the burden of homework and unnecessary curriculum rigor for test preparation, and to increase the opportunities for students to be active in the external communities (MOE, 2001).

In China, how to use the information of a variety of leadership models to build a meaningful and practical knowledge-base for Chinese school principals is a challenge (Chu, 2003). 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, Wang & Lin, 2006; Zhe, 2004; Zhe & 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. As long as the city-wide and nationwide entrance examination system is intact, much of the educational reform initiatives will be mission-impossible for schools and principals to accomplish.

This dissertation study is as much about the actual fit of the LCL framework and assessment as about seeking suitable pathways to compare and therefore benefit from educational theories and models cross-culturally. China is a huge country with immense diversities and disparities, in terms of its peoples, geographical areas, cultural practices, religious beliefs, and socio-economic development. Educational development is likely to be different in different parts of the country. A uniform educational policy is often not applicable across the country; local adaptation and modification are necessary. Politics and decisions regarding major policies have so far been largely confined to the State and party hierarchies (Pye, 1999), even though there is now some initial experimentation to allow local election of government officials by the people and to allow non-threatening, non-state social and political groups (Burns, 1999). People's input into policy has to go through government-controlled people's congresses at various levels. Encouragingly, despite the loss of some popular legitimacy, the government has so far demonstrated its ability to adapt to changes and maintain its grip on power (Schoenhals, 1999). The educational reform initiatives have been well financed and supported by the Ministry of Education. These features of Chinese politics have several implications for education where the State and party hierarchies still maintain unchallenged power for setting education policy. Conflicts and power struggle within the State and Party hierarchies often lead to abrupt changes and even reversals in educational policies. However, it is noticed that popular pressure for educational change has some possibility of being accommodated as long as it is not a threat to political stability and the party's power (Tsang, 2002).

Hard questions that pinpoint the issues that are 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?

Regardless how these questions will be answered, the promise of taking contextual factors into consideration needs to go beyond individual characteristics of the principal or micro-level indicators such as school size and school type. More importantly, macro-level political and social changes will inevitably influence the results of leadership assessment and how such results should be utilized. Ultimately, evaluation of the micro and macro context in which any assessment system resides might be necessary to achieve the best alignment of standards that lead to the success of schools and students and the instrument content.

The pendulum swings of educational reform efforts in both countries are striving to achieve the common objective of improved student learning. As illustrated in Figure 22, 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 United States, the reauthorization of NCLB is facing 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; in China, the push for "quality-oriented" education and curriculum reform has also been met with mounting difficulties in measuring intangible successes in student character-building, decreased efficiency of classroom activities, and enlarged gap between schools of high and low teacher capacities, just some of the similar issues faced by the U.S. educators (Chu & Cravens, 2008).

Balanced Education is the Common Goal

The U.S. Education

Increasing Standards
For Academic Learning
Without Sacrifizing
Social Learning

The Chinese Education

Improving Social
Learning Without
Reducing High Standards
for Academic Learning

The Balance Point
Academic and Social Learning
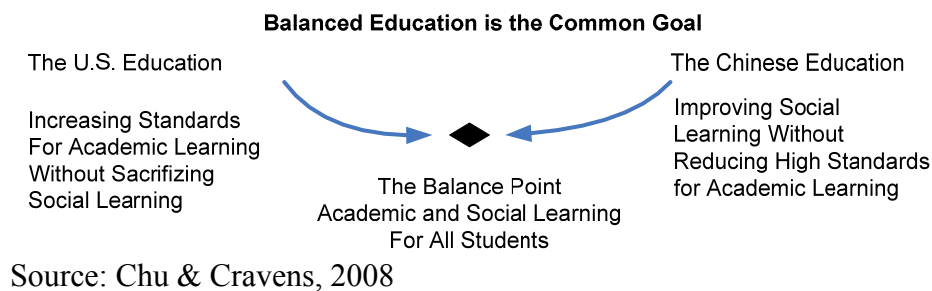For All Students

Source: Chu & Cravens, 2008

Figure 22: Seeking Balanced Education

The results of this dissertation will provide valuable 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 Learning-Centered Leadership 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. It is foreseeable that once identified and further validated in

the local context, such different and additional aspects of the construct can be incorporated into the existing applications of theories, e.g., the VAL-ED instrument, which will maximally predict outcomes for effective schools.

Without approaches that look at cultural, political and other factors that form the context of education in which schools and their players operate, the inferences and conclusions will not be valid. 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 dissertation converge and become meaningful.

APPENDIX

Appendix A: Learning-Centered Leadership Framework

Definitions of Core Components and Key Processes

(Porter, A., Goldring, E., Murphy, J., Elliott, S. N., & Cravens, X., 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 rigorous student academic and social learning.

- Rigorous 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 & 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.

Appendix B-a: 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: _____

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.												
Supporting	30. supports teachers' opportunities to improve their instructional practices.												

Appendix B-b: 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.
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.

Appendix B-c: The 4-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 <u>one</u> 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 C

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	Planning
Rigorous curriculum (content)	Implementing
Quality instruction (pedagogy)	Supporting
Culture of learning and professional	Advocating
Connections to external communities	Communicating
Systemic performance accountability	Monitoring
Your suggestions for additional components:	Your suggestions for additional processes:

Appendix D: Cognitive Interview Protocol

Vanderbilt University is studying the relevance and utility of a new leadership assessment to measure the leadership behaviors of school principals. In this portion of the study, we are testing various elements of the assessment so that we can determine the properties of the assessment and create a more complete, accurate test.

This assessment tool will be used to evaluate educational leadership. Assessment of each principal will be performed by the supervisor of the principal, the principal him or herself, and by all of the teachers of that school. Thus, we have asked you to participate in this interview as a representative of the type of individual that will be using this instrument.

This task asks you to read the survey aloud from the very beginning, including providing your name, title, etc. Please carefully read aloud all of the directions and comment on anything that seems unclear or that you might change. I may ask you a few specific questions related to some of the wording.

After reading the directions we ask that you begin to conduct the evaluation as if you were using it to rate the principal. Please continue to read and think aloud at all times. Again, I may ask you specific questions related to certain items or aspects of the survey.

*** *Special note for teachers:* The information that we obtain from this interview will only be used for research purposes. The actual document that you use to rate the principal will be destroyed and the recorded information will be held by only for purposes of recording notes relevant to the survey, but with no identifying information. Absolute confidentiality will be maintained at all times.

Since people are not used to thinking aloud, I'd like to show you an example of what I mean. I'm going to ask you to think out loud as you answer the question. Let me show you what I mean. (Turn to the example).

Cognitive Interview questions About the Form

Direction/Overall Opinions

- The instructions say "the principal may not have performed this behavior, but he or she ensured that it was performed." What does this mean to you?

***Re-write for principal:

- The instructions say "you may not have performed this behavior yourself, but you have ensured that it was performed." What does this mean to you?

Stem

- What does the stem, "The principal ensures the school …" mean to you?

Evidence

- How do you interpret the sources of evidence?

Effectiveness Rating

- How are you thinking about the choices for "effectiveness" on the rating scale?

Item-Level Prompts

Non-specific prompts to help the respondent think aloud:

- OK, I see, uh-huh, etc.

- Remember to think aloud as you answer the question.

- You're doing a great job thinking aloud.

- When you think out loud, it really helps us to understand how others approach these questions.

- (If respondent struggles with answer) Pretend this is a mail survey questionnaire you received in school, what would you do?

Prompts used to gain a better understanding of unclear items:

Can you say this question in your own words?

- What do you think this item is asking/measuring?

- OK, so what was your reasoning on that again? I just want to make sure I understand.

Prompts related to "sources of evidence":

- *If they are struggling to choose a source of evidence* "Does the evidence you are thinking of not fit any of our categories? If not, what is it?"

- Are there any sources of evidence that are missing from our choices? *Ask this when they are answering the question*

- *If they check "other sources" ask* "what other sources do you have in mind?"

- Would the item have been easier to rate if the effectiveness rating was before sources of evidence?

*** Each source of evidence mentioned should be followed by the following prompt at least *twice* through the course of the survey

- "You marked _____ as a source of evidence. Can you give an example of the evidence you had in mind?

Prompts related to "effectiveness":

*** Each score for the rating scale should be followed by the following prompt at least *twice* through the course of the survey

- "You marked _____ as the level of effectiveness. Can you give an example of what that rating means to you?

End of Survey Questions

- How much time is reasonable to ask teachers for willing participation in filling out this survey?

- How do you think teachers will react to filling this out?

- When you think about assessing principal leadership, are there other things that you think should be included?

- Do you have any concerns about this assessment?

REFERENCES

Adair, J. (1983). *Effective leadership*. Hants, England: Gower Publishing.

AERA, APA, & NCME. (1999). *The standards for educational and psychological testing*. Washington, DC: Authors.

American Educational Research Association, A. P. A., & National Council on Measurement in Education,. (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association

Angus, L. (1989). New leadership and the possibility of education reform. In J. Smyth (Ed.), *Critical perspective on educational Leadership* (pp. 63-92). London: Falmer.

Argenti, J. (1976). *Corporate collapse: The causes and symptoms*. London: McGraw-Hill.

Australian Principals Association Professional Development Council. (1994). *A Professional visits program for principals*

Bai, L. (2006). The learning-centered transition in leadership styles *Theory and Practice of Education 26*(5), 43-45.

Ball, S. (1998). Big Policies/Small World: an introduction to international perspectives in education policy. *Comparative Education, 34*(2), 119-130

Beck, L. G., & Murphy, J. (1994). *Ethics in educational leadership programs: An expanding role*. Newbury Park, CA: Corwin.

Blunt, P., & Jones, M. L. (1997). Exploring the limits of western leadership theory in East Asia and Africa. *Personnel Review 26*(1/2), 6-23.

Broadbent, J., Jacobs, K., & Laughlin, R. (1999). Comparing schools in the U.K. and New Zealand: Individualizing and socializing accountabilities and some implications for management control. *Management Accounting Research 10*, 339-361.

Brown, P., & Lauder, H. (1996). Education, globalisation and economic development. *Journal of Education Policy, 11*, 1-25.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-component analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* Washington, DC: American Psychological Association

Burns, J. (1999). The People's Republic of China at 50: National political reform. *China Quarterly, September*(159), 580-594.

Burroughs, S., Groce, E., & Webeck, M. L. (2005). Social studies education in the age of testing and accountability *Educational Measurement: Issues and Practice, 24*(3), 13-20.

Bush, T. (1998). The national professional qualification for headship: The key to effective school leadership? *School Leadership & Management, 18*(3), 321.

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

Chen, X. (2004, April 27). Diversification: The changing trend for the role of school principal *China Education Daily*

Chen, Z. (2006). Distributed leadership and school management transformation *Overseas Collection 69*(04).

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

Christy, W. K., & McNeil, L. (2000). *Excellence in administrator preparation programs: How will we know?* (142 Reports--Evaluative; 150 Speeches/Meeting Papers). Arkansas.

Chu, H. (2003). Principal professionalization: Knowledge base and support systems *Educational Administration Review, 1*, 223-252.

Chu, H. (2005). The current state of school management in Beijing: From the perspective of teachers. *??*

Chu, H. (2006). *Educational development in developing countries* Paper presented at the The International Seminar on Macro Education Policy in the 21st Century, National Center for Education Development Research, Beijing, China.

Chu, H. (2007). Education equity and efficiency: The dual goals of education reform. *Educational Development Review 12*.

Chu, H., & Cravens, X. C. (2008). Challenges for Chinese and U.S. principals: A comparative view of educational reform Beijing Normal University, Vanderbilt University.

Clark, D. L., Lotto, L. S., & Astuto, T. (1989). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly, 20* 41-68.

Cohen, Jacob (1960), A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* Vol.20, No.1, pp.37-46.

Coleman, J. S. (1990). *Equality and achievement in education*. Boulder, Colorado: Westview Press.

Collins, R. (1979). *The credential society*. New York: Academic Press

Conley, D. T., & Goldman, P. (1994). Ten propositions for facilitative leadership. In J. Murphy & K. S. Louis (Eds.), *Reshaping the principalship: Insights from transformational reform efforts*. Thousand Oaks, CA: Corwin Press.

Council of Chief State School Officers. (2002). *State standards and state assessment systems: A guide to alignment*.

Council of Chief State School Officers. (1996). *Interstate school leaders licensure consortium standards for school leaders*. Washington, DC: Author.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Darling-Hammond, L., Davis, S., L., LaPointe, M., & Myerson, D. (2005). *Developing successful school leaders*. New York: Wallace Foundation.

Dernberger, R. (1999). The People's Republic of China at 50: The economy. *The China Quarterly, 159*, 606-615.

Desimone, L. M., & Smith, T. M. y. v. i. p.-. (2005). Beyond accountability and average mathematics scores: Relating state education policy attributes to cognitive achievement domains *Educational Measurement, Issues and Practice, 24*(32), 5-.

DiMaggio, P. J., & Powell, w. W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields *American Sociological Review, 48*(2), 147-160.

Dimmock, C., & Walker, A. (1998). Transforming Hong Kong's schools: Trends and emerging issues. *Journal of Education Administration, 36*(5), 476-491.

Dimmock, C., & Walker, A. (2000). Developing comparative and international educational leadership and management: A cross-cultural model. *School Leadership & Management, 20*(2), 143.

Dorfman, P. W., & Howell, J. P. (1997). Leadership in western and Asian countries: Commonalities and differences in effective leadership processes across cultures. *Leadership Quarterly 8*(3).

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

Downing, S., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ Lawrence Erlbaum Assoc Inc.

Du, P. (2004). Highly effective principal leadership as the important condition for curriculum reform *Research on Educational Development 11*, 28-31.

Dubrin, A. J. (2004). *Leadership: Research findings, practice, and skills*. New York: Houghton Mifflin.

Ellett, C. D. (1999). Developments in the preparation and licensing of school leaders: The work of the Interstate school leaders licensure consortium. *Journal of Personnel Evaluation in Education, 13*(3), 201-2004.

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.

Elmore, R. F. (1995). Structural reform and educational practice *Educational Researcher, 24*(9), 23-26.

Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.

Elmore, R. F. (2007). Education: a "profession" in search of a practice. *Eye on Education by American Educational Research Association 15*(1-4).

Fairbank, J. K. (1994). *China, a new history*. Cambridge, MA: Belknap

Fang, X. (2005). Distributed leadership: Further exploring Western school leadership *Studies in Foreign Education, 12*(32), 57-60.

Fullan, M. (1982). *The meaning of educational change*. New York: NY: Teachers College Press.

Gao, R., Wang, B., & Lin, W. (2006). The foundation of moral theory - Research and reflection on leadership theory in China *Science and Technology Management Research 6*, 142-147.

Ginsberg, R., & Berry, B. (1990). The folklore of principal evaluation *Journal of Personnel Evaluation in Education 3*, 205-230.

Ginsberg, R., & Thompson, T. (1992). Dilemmas and solutions regarding principal evaluation *Peabody Journal of Education 68*(1), 58-74.

Glasman, N., & Heck, R. (1992). The changing leadership role of the principal: Implications for principal assessment. *Peabody Journal of Education 68*(1), 5-24.

Glasman, N. S. (1984). Student achievement and the principal. *Educational Analysis and Policy Analysis, 6*(3), 283-297.

Glasman, N. S., & Heck, R. H. (1992). The changing leadership role of the principal: Implications for principal assessment. *Peabody Journal of Education, 68*(1), 5-24.

Goldring, E., Cravens, X. C., Murphy, J., Elliott, S. N., Carson, B., & Porter, A. C. (2008). The Evaluation of Principals: What and How do States and Districts Assess Leadership? *The Elementary School Journal, Accepted*.

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

Goldring, E., & Vye, N. (2004). We must model how we teach: Learning to lead with compelling models of professional development. In W. Hoy & C. Miskel (Eds.), *Educational Administration, Policy, and Reform: Research and Measurement*. Greenwich, Connecticut: Information Age Publishing.

Green, T. F. (1985). *Excellence, equity, and equality*. New York, NY: Longman, Inc.

Gresham, E. M., Sugai, G., Horner, R. H., & Quinn, M. M. (1998). *Classroom and school-wide practices that support student's social competence: A synthesis of research* Washington, DC: Office of Special Education Programs.

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing, 20*(2), 225-240.

Guthrie, J. W., & Wong, K. K. (2007). Education finance from the perspective of politics, political cultures and government. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy*. New York, NY: Routledge.

Hair, J. F., Anderson, R. E., Tatham, R. l., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hallinger, P. (1992). Changing norms of principal leadership in the United States. *Journal of Educational Administration, 30*(3), 35-48.

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

Hallinger, P., & Kantamara, P. (2000). Educational change in Thailand: Opening a window onto leadership as a cultural process. *School Leadership & Management, 20*(2), 189.

Hallinger, P., & Leithwood, K. (1996). Culture and educational administration: A case of finding out you don't know you don't know *Journal of Education Administration 34*(5), 98-116.

Hallinger, P., & Leithwood, K. (1998). Unseen forces: The impact of social culture on school leadership *Peabody Journal of Education, 73*(2), 126-151.

Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal, 86*(2), 217-247.

Hallinger, P., & Murphy, J. (1986). The social context of effective schools *American Journal of Education 94*(3), 329-355.

Hallinger, P., & Murphy, J. (1987). Instructional leadership in the school context. In W. Greenfiled (Ed.), *Instructional leadership: Concepts, issues, and controversies*. Boston: Allyn & Bacon

Hallinger, P., Taraseina, P., & Miller, J. (1994). Assessing the instructional leadership of secondary school principals in Thailand. *School Effectiveness and School Improvement, 5*(4), 321-348.

Hambleton, R. (2002). Adapting achievement tests into multiple languages for international assessments. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 58-79). Washington, DC: National Academy Press.

Hambleton, R., Merenda, P., & Spielberger, C. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, N.J. : Lawrence. Erlbaum Associates.

Harkness, J. (1998). *Cross-cultural equivalence*. Mannheim, Germany: ZUMA.

Harlow, L. L. (2005). *The essence of multivariate thinking basic themes and methods* Mahwah, NJ Lawrence Erlbaum Assoc Inc.

Harris, B. M., & Monk, B. J. (1992). *Personnel administration in education* (3rd ed.). Needham Heights, MA: Allyn and Bacon.

Hart, A. W. (1992). The social and organizational influence of principals: Evaluating principals in context. *Peabody Journal of Education, 68*(1), 37-57.

Heck, R. H. (1996). Leadership and culture: Conceptual and methodological issues in comparing models across cultural settings. *Journal of Educational Administration, 34*(5), 74-97.

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

Heck, R. H., & Marcoulides, G. A. (1992). Principal assessment: Conceptual problem, methodological problem, or both? *Peabody Journal of Education, 68*(1), 124-144.

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

Heyneman, S. P. (2001). The growing international commercial market for educational goods and services *International Journal of Education Development 21*, 345-359.

Hu, Z. (2006). *Assessment and evaluation in education* (Second ed.). Guangzhou, China Guangdong Higher Education Publishing House.

Huang, W. (2004). Principal: From the role of an administrative position to the role of a management profession. *Educational Administration Review, 2*, 289-300.

Huff, J. (2006). *Measuring a leader's practice: Past efforts and present opportunities to capture what educational leaders do.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, CA: Sage.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrics 23*(187-201).

Kizlik, B. (2008). Measurement, assessment, and evaluation in education.   Retrieved April 11, 2008, from http://www.adprima.com/measurement.htm

Knapp, M. S., A., C. M., & Talbert, J. E. (2003). *Leading for learning: Reflective tools for school and district leaders (research report).* Seattle, WA: Center for the Study of Teaching and Policy.

Kochan, F. K., Bredeson, P., & Riehl, C. (2002). Rethinking the professional development of school leaders. In J. Murphy (Ed.), *The educational leadership challenge: Redefining leadership for the 21st century* (pp. 289-311). Chicago, Illinois: The National Society for the Study of Education.

Lambert, R. D. (2001). Updating the foreign language agenda. *The Modern Language Journal, 85*(3), 347.

Landis, J.R. and Koch, G. G. (1977) "The measurement of observer agreement for categorical data" in *Biometrics.* Vol. 33, pp. 159--174

Larson, M. S. (1977). *The rise of professionalism: A sociological analysis* Berkeley: University of California Press.

Lashway, L. (2003). *Improving principal evaluation. ERIC Digest. Access ERIC: FullText* (071 Information Analyses--ERIC IAPs No. EDO-EA-03-09). Oregon: ERIC Clearinghouse on Educational Management, Eugene, OR.

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

Leithwood, K. (2001). School Leadership in the context of accountability Policies *International Journal of Leadership in Education, 4*(3), 217-235.

Leithwood, K., Begley, P., & Cousins, B. (1992). *Developing expert leaders for future schools.* Bristol, PA: Falmer.

Leithwood, K., & Jantzi, D. (1993). Transformational leadership: How principals can help reform school cultures. *School Effectiveness and School Improvement, 1*, 249-280.

Leithwood, K., & Jantzi, D. (1999a). The relative effect of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly, 35*(5), 679-706.

Leithwood, K., & Jantzi, D. (1999b). Transformational school leadership effects: A replication. *School Effectiveness and School Improvement, 10*(4), 451-479.

Leithwood, K., & Jantzi, D. (2005). *A review of transformational school leadership research 1996-2005.* Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Leithwood, K. A., & Montgomery., D. J. (1986). *Improving principal effectiveness : The principal profile*. Toronto, Ontario: Ontario Institute for Studies in Education.

Li, S., & Zhao, R. (1999). A restudy of income distribution among Chinese inhabitants. *Economic Research Journal, April* 3-17.

Li, Y. (2003). Principal training needs analysis *Educational Administration Review, 1*, 279-296.

Li, Y. (2004). Reorientation for principals' role and function *Educational Administration Review 2* 301-312.

Lin, Q. (2003). *The study on the current Chinese implicit leadership theory: Investigating in Shenzhen, Guangzhou and Hanzhou* Jinan University Guangzhou

Lin, W., Chen, L., & Wang, D. (1987). Leadership behavior evaluation scale structure. *Psychology Research 2*, 199-207.

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

Ma, H., & Sun, X. (2006). A critical review of the suitability of Sergiovanni's moral education theory in China *Journal of Daqing Normal University 3*(3), 4-8.

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

Marcoux, J., Brown, G., Irby, B. J., & Lara-Alecio, R. A case study on the use of portfolios in principal evaluation. In (pp. 21): U.S.; Texas; 2003-04-00.

Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly, 39*, 370-397.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

Maxcy, S. J. (1991). *Educational leadership: A critical pragmatic perspective*. Toronto: OISE Press.

Ministry of Education. (1992). The draft opinions on enhancing the development of school principals in the nation.

Mintrop, H., & Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools. *Educational Evaluation and Policy Analysis, 29*(4), 319-352.

Mirvis, P., Ayas, K., & Roth, G. (2003). *To the desert and back: The story of the most dramatic business transformation on record*. San Francisco: Jossey-Bass.

Murphy, J. (1990). Principal instructional leadership. In L. L. L. P. W. Thurston (Ed.), *Advances in educational administration: Changing perspectives on the school* (Vol. 1, pp. 163-200). Greenwich, CT: JAI Press.

Murphy, J. (1991). *Restructuring schools: Capturing and assessing the phenomenon*. New York: Teachers College Press.

Murphy, J. (2005). *Connecting teacher leadership and school improvement*. Thousand Oaks, CA: Corwin

Murphy, J., Beck, L., Crawford, M., Hodges, A., & McGaughy, C. L. (2001). *The productive high school: Creating personalized academic communities*. Thousand Oaks, CA: Corwin.

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

Murphy, J., Goldring, E., Elliott, S. N., & Porter, A. (2006). *Learning-centered leadership: A conceptual foundation* New York, NY: Wallace Foundation

Murphy, J., & Hallinger, P. (1988). The characteristics of instructionally effective school districts. *Journal of Educational Research, 81*(3), 176-181.

National Association of Elementary School Principals. (2008). *Leading learning communities: Standards for what principals should know and be able to do* Washington, DC: Author.

National College of School Leadership. (2007). School self-evaluation: A reflection and planning guide for school leaders. from http://www.ncsl.org.uk/media/F08/88/self-evaluation-a-reflection-and-planning-guide.pdf

National Public Radio. (2007). Rookie Chicago principal faces early challenges [Electronic Version]. *Morning Edition*

Oyinlade, A. O. (2006). A method of assessing leadership effectiveness: Introducing the essential behavioral leadership qualities approach. *Performance Improvement Quarterly 19*(1), 25-40.

Peng, H. (2006). New development in school leadership: From instructional leadership to transformational leadership *Studies in Foreign Education, 2*(33), 12-16.

Porter, A., Goldring, E., Murphy, J., Elliott, S. N., & Cravens, X. (2006). *A conceptual framework for the assessment of principal and team school leadership*. New York, NY: Wallace Foundation

Porter, A. C., Polikoff, M. S., Goldring, E., Murphy, J., Elliott, S. N., & May, H. (2008). *Building a psychometrically sound assessment of school leadership: The VAL-ED as a case study*. Paper presented at the Annual Conference of American Educational Research Association New York, NY.

Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education* New York, NY: Wallace Foundation

Press, P. s. (1985). Documents on the reform of the educational system pp. 1-18.

Provus, M. (1971). *After discrepancy evaluation.* Paper presented at the Conference of the Association for Supervision and Curriculum Development, St. Louis, MO.

Pye, L. (1999). An overview of 50 years of the People's Republic of China: Some progress, but big problems remain. *The China Quarterly, September* (159), 569-579.

Qiao, J. (2003). Job analysis of principals in primary and secondary schools *Educational Administration Review, 1*, 253-278.

Rallis, S. F., & Goldring, E. B. (1993). Beyond the individual assessment of principals: School-based accountability in dynamic schools. *Peabody Journal of Education, 68*(2), 3-23.

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

Riehl, C. J. (2000). The principal's role in creating inclusive schools for diverse students: a review of normative, empirical, and critical literature on practice of education administration. *Review of Educational Research, 70*(1), 55-81.

Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the achievement gap*. Washington, DC: Economic Policy Institute.

Rowan, B., Raudenbush, S., & Kan, S. (1991). Organizational design in high schools: A multilevel analysis. *American Journal of Education, 99*(2), 238-266.

Schechter, C. (2008). Organizational learning mechanisms: The meaning, measure, and implications for school improvement. *Educational Administration Quarterly, 44*(2), 155-186.

Schoenhals, M. (1999). Political movements, change and stability: The Chinese communist party in power. *The China Quarterly, September* (159), 595-605.

Schwille, J. (1982). *Teachers as policy brokers in the content of elementary school mathematics*. East Lansing, Michigan: Institute for Research on Teaching, Michigan State University.

Scriven, M. (1996). Types of evaluation and types of evaluator *American Journal of Evaluation, 17*(2), 151-161.

Sebring, P., & Bryk, A. (2000). School leadership and the bottom line in Chicago. *Phi Delta Kappan, 81*(440-443).

Sergiovanni, T. J. (1992). *Moral leadership*. San Francisco: Jossey-Bass.

Silins, H. (1994). The relationship between transformational and transactional leadership and school improvement outcomes. *School Effectiveness and School Improvement, 5*, 272-298.

Sireci, S. G., Patsula, L., & Hambleton, R. K. (2005). Statistical methods for identifying flaws in the test adaptation process In R. Hambleton, P. Merenda & C. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, N.J. : Lawrence. Erlbaum Associates.

Spillane, J. P. (2006). *Distributed leadership*. San Francisco: Jossey-Bass.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Stine, D. O. (April 28, 2001). *Developing an evaluation system to improve principal performance and accountability*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, Washington.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research* Thousand Oaks, CA: SAGE Publications, Inc.

Sugrue, C. (2000). Changing leadership for changing times. *Journal of Education for Teaching, 26*(2), 189.

Synder, J., & Ebmeier, H. (1992). Empirical linkages among principal behaviors and intermediate outcomes: Implications for principal evaluation *Peabody Journal of Education, 68*(1), 75-107.

Thomas, D., Holdaway, E., & Ward, K. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education, 14*(3), 215-240.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association

Tsang, M. (1994). Costs of education in China: Issues of resource mobilization, equality, equity, and efficiency. *Education Economics, 2*(3), 287-312.

Tsang, M. (1996). The financial reform of basic education in China. *Education Review, 15*(4), 423-444.

Tsang, M. C. (2002). *School choice in People's Republic of China* (No. Occasional Paper No.9): Teachers College, Columbia University

van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*, 33-51.

van de Vijver, F. J. R., & Porrtinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. Hambleton, P. Merenda & C. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, N.J. : Lawrence. Erlbaum Associates.

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of Instructional practices on student achievement in science *Journal of Research in Science Teaching, 33*(10), 1110 - 1126.

Wang, J. (2004). Principal occupationization: The choice of educational reform *People's Education News 11*.

Wang, L., Chen, H., Yang, L., & Lin, X. (2004). The hierarchical structure of leadership personality characteristics. *Psychological Science, 27*(3), 677-681.

Wang, T. (2003). The systemic overview of principalship as an occupation. *Jiangsu College of Education Research Journal 3*.

Waters, T., & Grubb, S. (2004). *The leadership we need: Using research to strengthen the use of standards for administrator preparation and licensure programs*. Aurora, CO: Mid-continent Research for Education and Learning.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effects of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.

Wertsch, J. (1991). *Voices of the mind: A sociocultural approach to mediatedaction*. Cambridge, MA: Harvard University Press

Westbury, H. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? *Education Researcher 21*, 18-24.

Whyte, M. (1978). *The status of women in preindustrial society*. Princeton, NJ: Princeton University Press. .

Widenfelt, B. M. v., Treffers, P. D. A., Beurs, E. d., Siebelink, B. M., & 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, 8*(2), 135-148.

Wirt, F. M., & Kirst, M. W. (1982). *Schools in conflict : The politics of education* Berkeley, CA: McCutchan Pub. Corp.. .

Wong, K.-C. (1998). Culture and moral leadership in education *Peabody Journal of Education 73*(2), 106-125.

World Bank. (1998). *Regional disparities in China*. Washington, DC: East Asia Department, World Bank.

Yang, X. (2005). School leaders in transformational reforms. *Educational Research and Experiments 4*, 23-27.

Yukl, G. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Zhang, Z. (2005). The analysis of the role educational leaders and their strategies during the transition to formulate learning-centered schools *Studies in Foreign Education, 11*(32), 20-24.

Zhao, Q., & Wang, M. (2007). Construction of performance appraisal system for primary and secondary school principals-A case study in Beijing. *Educational Administration Review, 4*, 194-244.

Zhe, Y. (2006). Review of contemporary Western school leadership theories *Journal of Wenzhou Normal University, 25*(3).

Zhe, Y., & Li, C. (2006). Comparative theory and practice of school leadership management model *Journal of Wenzhou Normal University 27*(1), 113-116.