

Assessing the Measurement Properties of the Principal Instructional Management Rating Scale: A Meta-Analysis of Reliability Studies

Educational Administration Quarterly
49(2) 272-309

© The University Council for
Educational Administration 2013

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013161X12468149

eaq.sagepub.com



Phillip Hallinger¹, Wen-Chung Wang¹,
and Chia-Wen Chen¹

Abstract

Background: In a recent article, Hallinger (2011b) reviewed 135 empirical studies that had employed the Principal Instructional Management Rating Scale (PIMRS) over the prior three decades. The author concluded that the PIMRS appeared to have attained a consistent record of yielding reliable and valid data on principal instructional leadership. However, given the focus of the review, little detail was provided on measurement results related to the PIMRS instrument. **Research Design:** In light of the extensive past and current use of the PIMRS, the current article seeks to provide a comprehensive and detailed picture of reliability results. More specifically, the authors present a meta-analysis of reliability results derived from 52 data sets derived from 43 independent empirical studies in which the PIMRS had been employed for data collection. **Conclusions:** The main contribution of this article lies in elaborating on and updating information regarding the measurement properties of the PIMRS, among the most widely used scales in the

¹Hong Kong Institute of Education, Tai Po, Hong Kong

Corresponding Author:

Phillip Hallinger, Joseph Lau Chair Professor and Director of the Asia Pacific Centre for Leadership and Change at the Hong Kong Institute of Education, 10 Lo Ping Rd., Tai Po, Hong Kong, Hong Kong.

Email: hallinger@gmail.com

study and assessment of principal leadership. This information will be useful for researchers who are planning empirical studies of leadership and learning as well as practitioners who have a need for reliable measures of principal instructional leadership.

Keywords

principal, instructional leadership, PIMRS, reliability, meta-analysis

Research on school administrators for the period 1967-80 reminds one of the dictum: "The more things change, the more they remain the same." The state-of-the-art is scarcely different from what seemed to be in place 15 years ago. Although researchers apparently show a greater interest in outcomes than was the case in the earlier period, they continue their excessive reliance on survey designs, questionnaires of dubious reliability and validity, and relatively simplistic types of statistical analysis. Moreover these researchers persist in treating research problems in an ad hoc rather than a programmatic fashion. (Bridges, 1982, pp. 24-25)

Thirty years ago, Bridges (1982) highlighted persisting problems with the quality of research in educational leadership and management. In response to this and other critiques, researchers proposed stronger, research-based conceptual frameworks that sought to describe dimensions of the principal's instructional leadership role (e.g., Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger & Murphy, 1985; Leithwood & Stager, 1989; Ogawa & Bossert, 1995; Pitner, 1988). This conceptual progress was accompanied by the development of new research instruments (e.g., Hallinger & Murphy, 1985; Leithwood & Montgomery, 1986; Leithwood & Steinbach, 1991; Villanova, Gauthier, Proctor, & Shoemaker, 1981).

These research tools subsequently aided scholars in gaining a better understanding of the nature of the relationship between leadership and learning (e.g., Dwyer, 1986; Hallinger, Bickman, & Davis, 1996; Hallinger & Heck, 1996; Hallinger & Murphy, 1986; Heck, Larson, & Marcoulides, 1990; van de Grift, 1990). As a case in point, Hallinger (1983; Hallinger & Murphy, 1985) developed the Principal Instructional Management Rating Scale (PIMRS), a research tool that has been used in over 200 empirical studies conducted in 26 different countries (Hallinger, 2011b; Hallinger & Wang, 2013).¹

A recent review of research found that the PIMRS continues to be an instrument of choice among scholars studying principal leadership

(Hallinger, 2011b). The same article further suggested that the PIMRS has maintained a consistent record of yielding reliable and valid data. However, the earlier (Hallinger, 2011b) review was oriented toward a more general examination of research methodologies used in conjunction with the PIMRS and provided relatively few details concerning the reliability results obtained in this body of PIMRS studies. Given increasing global interest in instructional leadership and the continued widespread use of the PIMRS, the field will benefit from an updated and detailed assessment of its measurement properties. We note that there has not been any systematic attempt of this kind since the initial published report in the *Elementary School Journal* in 1985 (Hallinger & Murphy, 1985).

This report presents a meta-analysis of reliability results derived from 43 independently conducted studies that employed the PIMRS as a research tool over the past three decades. The study addresses several research questions:

Research Question 1: Does the PIMRS provide reliable data for the purposes of assessing principal instructional leadership in research and practice?

Research Question 2: How do reliability estimates differ based on the role group of the respondents (i.e., teachers or principals)?

Research Question 3: Does the PIMRS yield reliable data when used in rating principals working at different school levels (i.e., primary, middle, secondary) and cultural contexts (i.e., nations)?

The relevance of this information was inadvertently confirmed by a recent publication that critiqued the reliability of several leadership scales (Condon & Matthews, 2010). However, the information included on the PIMRS in the report was incomplete and out of date. This further highlights the need for a comprehensive, up-to-date description of the reliability of the PIMRS. This will aid scholars and practitioners in choosing among instruments used for assessing principal instructional leadership and in making methodological choices when using the PIMRS.

Theoretical Perspective

In this section of the article, we provide background information on the conceptual framework that informed development of the PIMRS. Then we discuss the development of the PIMRS instrument. This will set the stage for presentation of the methodology and results.

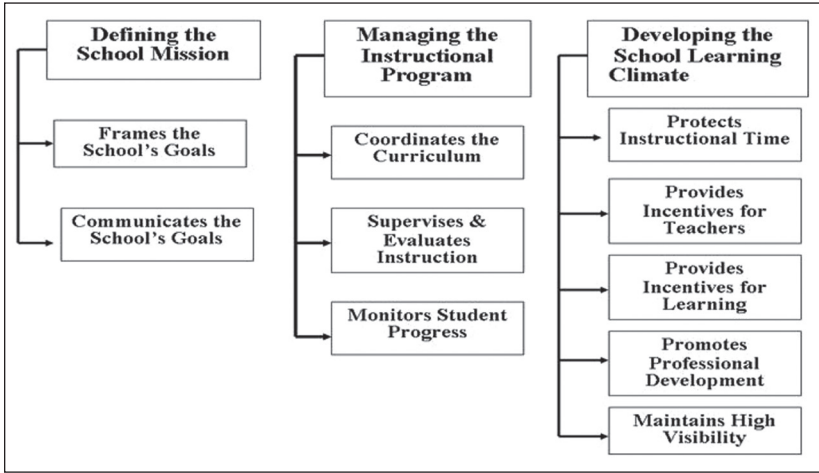


Figure 1. Principal Instructional Management Rating Scale (PIMRS) conceptual framework.

The PIMRS Framework

The PIMRS (Hallinger, 1982/1990) is grounded in a conceptual framework that proposes three dimensions in the instructional leadership role: Defines the School Mission, Manages the Instructional Program, and Develops a Positive School Learning Climate (Hallinger & Murphy, 1985; see Figure 1). These three dimensions are delineated into 10 instructional leadership functions. Two functions, frames the school’s goals and communicates the school’s goals, comprise the dimension Defines the School Mission. These functions concern the principal’s role in working with staff to ensure that the school has a clear mission and that the mission is focused on academic progress of its students. While this dimension does not assume that the principal defines the school’s mission alone, it does propose that the principal is responsible for ensuring that such a mission exists and for communicating it widely to staff.

The second dimension is Manages the Instructional Program. This incorporates three leadership functions: supervises and evaluates instruction, coordinates the curriculum, and monitors student progress. This dimension focuses on the role of the principal in “managing the technical core” of the school (Hallinger & Murphy, 1985). In today’s schools, it is clear that the principal is not the only person involved in monitoring and developing the school’s

To what extent does your principal. . . ?					
I. FRAME THE SCHOOL GOALS					
	Almost Never		Almost Always		
	1	2	3	4	5
1. Develop a focused set of annual school-wide goals	1	2	3	4	5
2. Frame the school's goals in terms of staff responsibilities for meeting them	1	2	3	4	5
3. Use needs assessment or other systematic methods to secure staff input on goal development	1	2	3	4	5
4. Use data on student academic performance when developing the school's academic goals	1	2	3	4	5
5. Develop goals that are easily translated into classroom objectives by teachers	1	2	3	4	5

From *PIMRS* - Teacher Form 1.3 (Hallinger, 1983, p. 2)

Figure 2. Sample Principal Instructional Management Rating Scale (PIMRS) rating subscale: teacher form.

instructional program (Hallinger, 2011a; Leithwood, Day, Sammons, Harris, & Hopkins, 2006). At the same time, however, coordination and control of the academic program of the school remains a key leadership responsibility of the principal, even when tasks are delegated or shared.

The third dimension, Develops a Positive School Learning Climate, includes several functions: protects instructional time, promotes professional development, maintains high visibility, provides incentives for teachers, and provides incentives for learning. This dimension is broader in scope than the other dimensions and overlaps with factors generally associated with transformational leadership (e.g., Hallinger, 2003; Leithwood et al., 2006; Marks & Printy, 2003). This dimension conforms to the notion that successful schools create an “academic press” through the development of high standards and expectations and a culture that fosters and rewards capacity development and continuous learning (Hallinger & Murphy, 1985).

Development of the PIMRS Instrument

The PIMRS instrument. The original form of the PIMRS (Hallinger, 1982/1990) contained 11 subscales and 72 “behaviorally anchored” items. Subsequent revision of the instrument reduced the instrument to 10 subscales and 50 items (Hallinger, 1982/1990, 1983).² For each item, the rater assesses the frequency with which the principal enacts a behavior or practice associated with the particular instructional leadership function. Each item is rated on a Likert scale ranging from (1) *almost never* to (5) *almost always* (see Figure 2).

The instrument is scored by calculating the mean for the items that comprise each subscale. This results in a data-based profile of principal performance.

Three parallel forms of the instrument have been developed and tested for completion by the principal, teachers, or a supervisor. The items that comprise each form are identical; only the stems change to reflect the differing perspectives of the role groups. Hallinger (2011b) noted that researchers consistently report significant differences between teacher and principal perceptions of the principal's instructional leadership. Moreover, principal self-report scores tend to be substantially higher than those obtained from teachers (e.g., Benoit, 1990; Brown, 1991; Corkill, 1994; Dennis, 2009; Haack, 1991; Haasl, 1989; Hallinger, 1983; Henderson, 2007; Krug, 1986; Mallory, 2002; Marshall, 2005; Meek, 1999; Meyer, 1990; O'Day, 1984; O'Donnell, 2002; Reid, 1989; Shatzer, 2009; S. Smith, 2007; Stevens, 1996; Vinson, 1997).

Notably these "role set" (Merton, 1957) differences in PIMRS ratings obtained from teachers and their principals extend to contexts other than the United States. Empirical comparisons have yielded a similar pattern of results in Thailand (Hallinger & Lee, in press; Poovatanikul, 1993; Ratchaneeladdajit, 1997; Taraseina, 1993), Guam (San Nicolas, 2003), the Philippines (Saavedra, 1987; Salvador, 1999; Yogere, 1996), the Maldives (Wafir, 2011), Hong Kong (Chan, 1993), and Taiwan (Chi, 1997; Tang, 1997; Yang, 1996). In addition, despite differences in the magnitude of ratings obtained by the two role groups, researchers often report a similar pattern in their ratings on the various subscales that comprise the PIMRS.

Five studies have been conducted that included extensive assessments of the reliability and validity of the PIMRS. Three were conducted in the United States (Hallinger, 1983; Howe, 1995; Jones, 1987), one in Thailand (Taraseina, 1993), and one in Cameroon (Wotany, 1999). Hallinger's (1983) study focused on the elementary school level and the others on secondary schools. Numerous additional studies have reported the results of reliability tests (see Tables 1 and 2). In the following sections, we provide an overview of the approaches employed in assessing the reliability and validity of the PIMRS instrument.

Assessing the reliability of the PIMRS. Lang and Heiss (1998) defined reliability as the consistency with which an instrument yields the same or similar responses across settings and time. Several different approaches may be employed for assessing the reliability of a test instrument: test-retest, parallel forms, and internal consistency (Kerlinger, 1966). Studies employing the PIMRS have relied exclusively on measures of internal consistency. Internal consistency refers to the degree to which items that have been grouped together conceptually as subscales correlate with each other (Kerlinger, 1966).

Table 1. Principal Data Sources.

Number	Author(s)	Year	Data	Nation	N (Principals)	School Level
1	Anderson	2006	Extracted	United States	190	Secondary
2	Babcock	1991	Extracted	United States	213	Primary
3	Carr	2011	Raw	United States	6	Primary
4	Carson	2012	Raw	United States	77	Secondary
5	Dunn	2010	Extracted	United States	128	Primary/ secondary
6	Gjelaj Merturi	2010	Raw	United States	10	Primary
7	Goldring, Cravens, Murphy, Porter, and Elliott	2012	Raw	United States	58	Primary/ secondary
8	Greb	2011	Raw	United States	31	Primary
9	Hallinger & Lee	In press	Raw	Thailand	1195	Primary/ secondary
10	Long	2008	Raw	United States	67	Secondary
11	Lyons	2010	Raw	United States	15	Secondary
12	Minus	2010	Extracted	United States	62	Middle
13	Munroe	2009	Raw	United States	35	Primary
14	Nogay	1995	Extracted	United States	61	Secondary
15	Peariso	2011	Extracted	United States	36	Secondary
16	Shafeeu	2011	Raw	Maldives	10	Primary/ secondary
17	Todd	2006	Raw	United States	122	Secondary
18	Wang	2011	Raw	Chin	23	Secondary
19	Wong	2010	Raw	Malayasia	195	Primary/ secondary

Table 2. Teacher Data Sources.

Number	Author(s)	Year	Data	Nation	N (Teachers)	N (Schools)	School Level	Reliability Test
1	Adam	2012	Raw	United States	128	9	Secondary	Generalizability theory
2	Aste	2009	Extracted	United States	248	6	Primary	Alpha
3	Carson	2012	Raw	United States	95	77	Secondary	Generalizability theory
4	Chappellear	2011	Extracted	United States	213	18	Secondary	Alpha
5	Clark	2009	Extracted	United States	356	90	Primary	Alpha
6	Coltharp	1996	Extracted	United States	128	14	Primary	Alpha
7	Dale	2010	Raw	United States	177	36	Primary	Generalizability theory
8	DuPont	2009	Extracted	United States	126	1	Primary/ secondary	Alpha
9	Fancera	2009	Raw	United States	580	100	Secondary	Generalizability theory
10	Fulton	2009	Extracted	United States	174	24	Secondary	Alpha
11	Gallon	1998	Extracted	United States	514	34	Primary/ secondary	Alpha
12	Geiselman	2004	Extracted	United States	979	180	Primary	Alpha
13	Goldring, Cravens, Murphy, Porter, and Elliott	2012	Raw	United States	1,610	58	Primary/ secondary	Generalizability theory
14	Greb	2011	Raw	United States	107	31	Primary	Generalizability theory
15	Hallinger	1983	Extracted	United States	104	10	Primary/ secondary	Ebel
16	Hart	2006	Extracted	International	524	18	Primary/ secondary	Alpha
17	Howe	1995	Extracted	United States	393	34	Primary	Ebel
18	Jones	1987	Extracted	United States	405	29	Secondary	Ebel
19	Leitner	1989	Extracted	United States	412	26	Primary	Ebel
20	Lyons	2010	Raw	United States	176	15	Secondary	Generalizability theory
21	MacNeil	1992	Extracted	United States	903	56	Primary	Alpha

(continued)

Table 2. (continued)

Number	Author(s)	Year	Data	Nation	N (Teachers)	N (Schools)	School Level	Reliability Test
22	Minus	2010	Extracted	United States	298	62	Middle	Alpha
23	Nogay	1995	Extracted	United States	424	76	Secondary	Alpha
24	O'Day	1984	Extracted	United States	140	19	Primary	Alpha
25	Ponussamy	2010	Raw	Malaysia	105	14	Primary	Generalizability theory
26	Roudebush	1996	Extracted	United States	358	12	Primary	Alpha
27	Shafeeu	2011	Raw	Malaysia	201	10	Primary/ secondary	Generalizability theory
28	Shatzer	2009	Raw	United States	280	37	Primary	Generalizability theory
29	Sheppard	1996	Extracted	Canada	624	50	Primary/ secondary	Alpha
30	Taraseina	1993	Extracted	Thailand	774	10	Secondary	Ebel
31	Wafir	2011	Extracted	Maldives	225	5	Primary	Alpha
32	Wang	2011	Raw	China	156	23	Secondary	Generalizability theory
33	Waters	2005	Extracted	United States	127	7	Primary	Alpha

PIMRS studies have varied with respect to the form of the scale that was used (i.e., teacher or principal), the level of the scale on which reliability was calculated (i.e., whole scale, dimension, function), as well as the school level (i.e., primary, middle, secondary) and cultural context in which the study was conducted.

Gathering data with the PIMRS directly from principals represents a type of self-assessment. The resulting score reflects a latent trait of the individual subject (Kerlinger, 1966). This is a typical case faced in measurement for which researchers often employ Cronbach's (1951) alpha test of internal consistency. In our data set, all PIMRS studies that reported reliability results based on principal self-report data employed Cronbach's alpha (see Tables 1 and 3).

Researchers have, up until now, employed two different statistical tests to assess the scale's internal consistency with data collected from teachers. Although the majority have applied Cronbach's alpha to data obtained with the PIMRS Teacher Form, several employed Ebel's (1951) test of reliability. For the purposes of the present background discussion of the PIMRS's reliability,

Table 3. Meta-Analysis of Reliability Estimates Derived From Principal Respondent Studies.

Number	Author(s)	Year	Data Source	Nation	School Level	N (Principals)	Whole Scale (Principals)	Create Mission Instruction	Manage Instruction	Develop Climate	Frame Goals	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Incentives for Professional Development	Protect Instructional Time	
1	Anderson	2006	Extracted	United States	Secondary	190	—	.90	.92	.87	—	—	—	—	—	—	—	—	—	
2	Babcock	1991	Extracted	United States	Primary	213	—	—	—	—	.82	.75	.69	.72	.64	.65	.66	.76	.68	.76
3	Carson	2012	Extracted	United States	Secondary	77	.91	.79	.87	.80	.68	.71	.73	.81	.75	.51	.47	.79	.68	.66
4	Dunn	2010	Extracted	United States	Primary/secondary	128	.94	—	—	—	.86	.86	.57	.88	.85	.79	.86	.86	.85	.60
5	Golding Cravens, Murphy, Porter, and Elliott	2012	Raw	United States	Primary/secondary	58	.96	.84	.93	.92	.79	.74	.83	.83	.85	.70	.77	.81	.75	.72
6	Greb	2011	Raw	United States	Primary	31	.85	.80	.83	.59	.80	.76	.88	.75	.65	.68	.65	.61	.67	.61
7	Hallinger & Lee	In press	Raw	Thailand	Primary/secondary	1,195	.96	.88	.91	.94	—	—	—	—	—	—	—	—	—	—
8	Long	2008	Raw	United States	Secondary	67	.99	.95	.97	.97	.92	.88	.93	.92	.92	.90	.92	.92	.91	.90
9	Lyons	2010	Raw	United States	Secondary	15	.91	.84	.88	.77	.64	.79	.24	.86	.82	.54	.70	.26	.68	.70
10	Minus	2010	Extracted	United States	Middle	62	—	—	—	—	.81	.77	.81	.71	.80	.75	.48	.73	.80	.75
11	Munroe	2009	Raw	United States	Primary	35	.93	.80	.88	.85	.63	.74	.77	.74	.81	.62	.66	.78	.74	.78
12	Nogay	1995	Extracted	United States	Secondary	61	.93	—	—	—	.86	.83	.60	.83	.72	.67	.74	.72	.71	.72
13	Peariso	2011	Extracted	United States	Secondary	36	—	—	—	—	.63	.62	.56	.67	.73	.58	.56	.62	.72	.79
14	Todd	2006	Raw	United States	Secondary	122	.97	.92	.94	.94	.93	.81	.88	.88	.87	.87	.51	.90	.86	.88
15	Wang	2011	Raw	China	Secondary	23	.98	.90	.91	.96	.88	.84	.85	.80	.92	.88	.82	.80	.92	.80
16	Wong	2010	Raw	Malaysia	Primary/secondary	195	.94	.83	.88	.90	.77	.74	.77	.77	.82	.82	.64	.77	.76	.84
Summary statistics						2,508	.96	.88	.91	.93	.85	.79	.80	.83	.82	.78	.74	.82	.80	.80

Note. These analyses include all data sets comprised of 15 or more principals. All calculations are based on Cronbach's alpha test of internal consistency. "Extracted" refers to alpha coefficients extracted from research reports. "Raw" refers to our analysis of actual data sets obtained from the authors of the studies.

we simply wish to highlight the fact that our data set contains reliability results that were obtained through a variety of methods (see Tables 1 and 2). While this presented challenges for our attempt to synthesize findings across studies, the diversity of approaches used to formulate and express the scale's reliability can also be viewed as a strength of the current study. We will discuss these points in greater detail in the methods section of the article.

Assessing the validity of the PIMRS. An appraisal instrument must not only provide accurate (i.e., reliable) data, but also measure the construct as conceptualized by the researcher (i.e., validity; Lang & Heiss, 1998). In the validation studies cited earlier, four criteria were used to judge the validity of the PIMRS:

1. Content validity: Items making up each subscale of the instrument must be relevant to the critical requirements of the job; each item assigned to a subscale must achieve a minimum average agreement of .80 among a group of raters.
2. Discriminant validity: The subscales should discriminate among principals; that is, the variance in principal ratings within schools must be less than the variance in ratings of principals between schools.
3. Construct validity (subscale intercorrelation): Groups of items within a subscale must intercorrelate more strongly with each other than with other subscales.
4. Construct validity (documentary support): An analysis of school documents related to the instructional management behavior of the principals should yield profiles of the principals' instructional management performance similar to those obtained from teachers on the questionnaire.

The application of these methods to development of the PIMRS yielded an instrument that met common standards of internal validity in the original validation study conducted by Hallinger (1983). Studies by Jones (1987), Howe (1995), and Taraseina (1993) largely replicated the original validation study's results at the secondary school level in studies conducted in Canada, the United States, and Thailand, respectively. We note, however, that no formal studies have assessed the external validity of the scale. Given space limitations, we do not offer further description of validation procedures or results in this article. Readers interested in such detail can consult the original research reports (Hallinger, 1983; Hallinger & Murphy, 1985; Howe, 1995; Jones, 1987; Taraseina, 1993; Wotany, 1999) as well as a PIMRS Technical Report (Hallinger & Wang, 2013) which describes our recent effort to revalidate the scale.³

Method

In this section of the article, we clarify the nature of the data sources for this study, the types of reliability tests that have been applied to PIMRS data, and our approach to integrating reliability estimates obtained across the studies.

Data Sources

We were fortunate to gain access to secondary data contained in 43 previously conducted studies that had employed the PIMRS for data collection. The data consisted of two types. The first was statistical information contained in published research reports and doctoral dissertations. The second consisted of raw data sets obtained directly from researchers who had used the scale. We present these data sources separately since they presented different challenges and required different treatment in data analysis.

Extracted data. The research began with a literature search that sought to identify the full set of studies that had used the PIMRS. Our search found 15 master degree theses, 151 doctoral dissertations, and 11 published journal articles.⁴ The 177 studies⁵ were reviewed in order to identify ones that had conducted reliability analyses. We located 28 studies that reported reliability results (see Tables 1 and 2).

We proceeded to extract information from these studies. The data included the author(s), year, national context, school level(s), respondent group(s), sample size(s), reliability test, and reliability results. These data were entered into a table in MS Excel.

These research reports varied widely in terms of the sample size, role group of respondents, method of calculating reliability, and levels of the scale for which reliability was reported. As in all meta-analyses, variability in the design and methodology of the component studies presented challenges in terms of our goal of quantitative integration of the data. If, for example, the reader compares the upper and lower sections of Table 5, the limitations of relying on data extracted from published reports quickly become apparent. Lack of comparable data would have impeded our attempt to develop a comprehensive assessment of the PIMRS's reliability. For this reason, we sought to obtain original item-level data for our calculations.

Raw data sets. Given the frequent continuing use of the PIMRS instrument among researchers,⁶ it appeared feasible to try and obtain raw data from recently completed studies. Raw item-level data would offer several advantages for meta-analysis. First, it allows researchers to use a consistent test for calculating reliability estimates. Second, it affords the opportunity to

compare the pattern of reliability estimates obtained from different methods of calculation. Third, access to item-level data enables calculation of a more comprehensive set of reliability estimates (e.g., full scale, three dimensions, and 10 functions). In contrast, the reliability estimates extracted from research reports typically focused only on one, or at most two, of the three scale levels. In sum, access to raw data enables researchers to exploit the full power of meta-analytic techniques, resulting in a more robust synthesis of results.

We subsequently contacted authors of PIMRS studies that had been completed since 2008 with the goal of gaining access to original item-level PIMRS data. We were able to obtain access to 25 original data sets comprised of PIMRS responses from 19 different researchers.⁷ When these were combined with the extracted data, we had 52 data sets derived from 43 independent studies. The data sets were separated into two groups, based upon the data source, either principals (see Table 1) or teachers (see Table 2). We describe these in turn.

Principal self-report data. Principals represented the data source in 19 of the studies (see Table 1). As shown in Table 1, this consisted of 13 raw data sets and 6 data sets comprised of extracted data. We eliminated 3 studies that had surveyed fewer than 15 principals from our analyses (i.e., Carr, 2012; Gjelaj Merturi, 2010; Shafeeu, 2011) on the grounds that this was the minimum level needed in order to obtain sound results. This left us with 16 studies in the principal respondent data set. Four had collected data in East Asia and 12 in the United States (see Table 3). The studies were distributed across all school levels (i.e., primary, middle, secondary schools).⁸ The sample size of the studies ranged from 15 to 1,195 principals, with a mean of 157 principals per study and a total sample of 2,508 principals.

Given the large size and uniqueness of the Hallinger and Lee (in press) data set from Thailand, elaboration on its characteristics seems appropriate. Unlike the other studies, this research employed a shortened version of the principal form of the PIMRS comprised of 20 items. This short form of the instrument was used to collect data from a nationally representative sample of 1,195 principals in Thailand. In order to prepare these data for meta-analysis, we applied a set of procedures to transform the 20-item scale so that the reliability results would approximate those of the 50-item test.

We employed the Spearman Brown test to correlate the items included in this study with results obtained from the raw data sets that had employed the full PIMRS comprised of 50 items. This procedure resulted in revised reliability estimates for both the full PIMRS as well as its three dimensions. The distribution of items in the short form was, however, insufficient to obtain an accurate estimation of reliability for the 10 leadership functions.

Teacher data sources. The teacher respondent data set consisted of data gathered in 33 studies conducted between 1983 and 2012 (see Table 2). The 33 studies were comprised of 11 raw data sets and 22 extracted data sets (see Table 2). The studies came from six different countries: United States (25), Canada (2), China (1), Thailand (1), Malaysia (2), the Maldives (1), as well as one study that analyzed a cross-cultural data set (Hart, 2006). The sample size of these teacher data sets ranged from 95 to 1,610, with a mean of 359 teachers per study and a total sample of 10,080 teachers. Respondents included teachers in primary, middle, and secondary schools.

Data Analysis

In this section we begin by describing the procedures for analyzing the reliability of the principal and teacher data. Then we discuss the meta-analytic procedures used to mathematically integrate results obtained across the studies. Finally, we discuss the standards for assessing the results of reliability analyses.

Procedures for assessing the reliability of the PIMRS Principal Form. As noted earlier, it was standard procedure for the researchers to employ Cronbach's alpha in testing the reliability of the PIMRS Principal Form. First, we extracted alpha reliability estimates from the research reports. Then we calculated Cronbach's alpha coefficients for the raw data sets. Finally, we combined the full set of alpha reliability coefficients into a single MS Excel table in preparation for meta-analysis.

Procedures for assessing reliability of the PIMRS Teacher Form. For reasons suggested earlier, synthesizing reliability estimates from the data sets comprised of teacher responses was less straightforward. Researchers applied different reliability tests to these data sets (e.g., see Tables 2 and 5). Although most researchers employed Cronbach's test, some researchers suggested that this application of coefficient alpha violates a fundamental assumption of the test (e.g., see Jones, 1987). When analyzing a PIMRS data set obtained from teacher respondents, Cronbach's test treats each teacher's response independently, as if each teacher were rating a different principal. In reality, however, teachers are nested within schools, with each school's teachers rating their own principal. In this case, the reliability estimates of internal consistency should be based on the combined ratings of teachers grouped together by their schools.

Thus, several researchers (e.g., Howe, 1995; Jones, 1987; Leitner, 1989; Taraseina, 1993) employed Ebel's (1951) test, which yields a reliability

rating based on the aggregated responses of teachers within their respective schools. The reliability result is therefore treated as a feature of the school (i.e., the principal). Ebel's formula is: $r = (\sigma_{\text{bet}}^2 - \sigma_{\text{wit}}^2) / \sigma_{\text{bet}}^2$, where r is the reliability of average ratings, σ_{bet}^2 is the between-groups variance, and σ_{wit}^2 is the within-groups variance. When employing this formula with the PIMRS, the researchers most frequently applied the test to each of the 10 leadership functions (e.g., see Tables 2 and 5).

When examining the various studies that had collected teacher data with the PIMRS, we found that 18 researchers had derived reliability estimates using Cronbach's alpha and four from Ebel's test. While both tests offer estimates of internal consistency, the coefficients cannot be combined through meta-analysis. Therefore, these data were grouped separately for subsequent meta-analyses (see Tables 5 and 6).

When we turned our attention to the raw data sets, we faced a decision as to which test to use. Although we initially considered Ebel's (1951) test superior to Cronbach's alpha for analyzing reliability of the teacher data, this approach method still had two problems (Schmitt, 1996). First, Ebel's formula assumes that teachers are randomly selected from the same population. This assumption is not, however, tenable because teachers are nested within schools and each school's teachers are rating their particular principal. The other problem is that item-level scores are "invisible" (i.e., ignored) in the Ebel formula, which only employs the total score from a teacher on the relevant subscale.

With these limitations in mind, we employed a formula based on generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Kane, Gilmore, & Crooks, 1976) for testing the reliability of raw data sets comprised of teacher responses. This approach not only takes into account the hierarchical structure of the data but also utilizes item-level scores. Thus, we assert that a more accurate way of calculating the reliability of scores obtained from teacher respondents is represented in the following formula.

$$\rho_p = \frac{\sigma_p^2}{\sigma_p^2 + \frac{1}{n_t n_i} \sigma_e^2}, \quad (1)$$

where σ_p^2 is the variance of principals, σ_e^2 is the error variance, n_t is the number of teachers per school, and n_i is the number of items. It should be noted that the number of teachers per school is assumed to be identical across schools in this formula. In cases where schools have different numbers of teachers, one can use their harmonic mean to substitute for n_t . The reliability for principal ρ_p can then be calculated as follows:

$$\hat{\rho}_p = \frac{[MS_p - MS_{p \times t} - MS_{p \times i} + MS_e]}{[MS_p - MS_{p \times t} - MS_{p \times i} + MS_e] + MS_e}, \quad (2)$$

where MS_p is the mean square of principal, $MS_{p \times t}$ is the mean square of principal by teacher, $MS_{p \times i}$ is the mean square of principal by item, and MS_e is the mean square of error (Kane et al., 1976).

We will refer to this approach as a “generalizability theory” or “gen theory” test.⁹ We used this formula to compute the reliability in studies where we had raw, item-level teacher response data. It was not, however, possible to reanalyze data extracted from published reports with this method due to lack of item-level information. Thus, we were left with three separate groups of teacher respondent studies, each based on reliability estimates obtained through different tests. We further note that although each test produced a set of reliability coefficients, the results of the three tests could not be combined in subsequent meta-analytic procedures. We report both the reliability results of individual studies and the summary statistics obtained through meta-analysis in separate portions of Table 5 and 6.

Meta-analytic procedures. Meta-analysis refers to statistical techniques that integrate findings obtained from a body of research comprised of independent studies (Glass, 1977). Meta-analysis is frequently applied as a means of understanding the trend in substantive findings across studies (Glass, 1977). The meta-analysis of leadership effects studies in education conducted by Robinson, Lloyd, and Rowe (2008) is an example of this species of meta-analysis.

However, Rodriguez and Maeda (2006) also applied meta-analytic methods to the generalization of reliability coefficients. This involves mathematically synthesizing the reliability coefficients obtained from different studies weighted by their sample sizes. The resulting generalization reliability estimate is more accurate than any single reliability coefficient (Rodriguez & Maeda, 2006).

Take the alpha coefficient as an example. It is assumed that each of K studies ($k = 1, \dots, K$) provides an estimate of the population alpha coefficient. Let α_k be the alpha coefficient, n_k be sample size, and J_k be the number of items in study k . The alpha coefficient should be transformed as:

$$T_k = (1 - \alpha_k)^{1/3}, \quad (3)$$

with error variance:

$$v_k = \frac{18J_k(n_k - 1)(1 - \alpha_k)^{2/3}}{(J_k - 1)(9n_k - 11)^2}. \quad (4)$$

Then, the weighted mean transformed alpha is:

$$\bar{T} = \frac{\sum(T_k / v_k)}{\sum(1/v_k)}, \quad (5)$$

with error variance of $1/\sum(1/v_k)$. Finally, the generalization reliability for the alpha coefficient can be obtained as $|1 - \bar{T}^3|$.

In sum, this approach to meta-analytic transformation of data provides a "weighted average" of the reliability estimates derived from a larger set of studies (Hakstian & Whalen, 1976). The weighted average adjusts for the sample size of the particular studies, giving greater weight to studies with larger samples. As noted previously, we categorized the data sets for the studies that contained teacher responses (see Tables 5 and 6) in order to maintain the integrity of the results of different reliability tests.

Standards for assessing scale reliability. Contrary to common belief, there is no single standard for assessing the reliability of a research instrument (Kerlinger, 1966; Latham & Wexley, 1981; Smith & Kendall, 1963). The appropriate standard should be based on the *intended use of the data*. In general, instruments used for research and assessments of job performance assessment (e.g., principal evaluation) are evaluated according to different standards (Latham & Wexley, 1981). For example, in situations where important decisions about the employment of individuals are made on the basis of test scores, Latham and Wexley (1981, p. 66) proposed a minimum standard of .80 for tests of reliability. Nunnally (1978) suggested that performance assessment instruments should achieve a reliability standard of at least .90, or even .95, when the data will be used for personnel decisions.

A lower standard is, however, generally applied when data are employed for research purposes. For example, Hair and colleagues suggested a minimum acceptable range of .60 to .70 when data will be used for research (Hair, Black, Babin, Anderson, & Tatham, 1998). Other scholars (e.g., Fraenkel & Wallen, 1990; Kerlinger, 1966; Lang & Heiss, 1998; Nunnally, 1978) have recommended that research instruments should meet a minimum reliability standard of .70. Nunnally (1978, p. 245) even went so far as to assert that efforts to achieve reliability levels beyond .80 are a waste of resources when instruments are intended for use in basic research.

It is also useful to note Gay's (1992) observation that reliability is influenced by the length of the instrument. Longer scales tend to yield greater reliability. Thus, the subscales that comprise an instrument typically yield lower reliability coefficients than the full scale. This is relevant for our study of the reliability of the PIMRS, which has been analyzed in terms of the full scale as well as its three dimensions and 10 functions.

These perspectives on standards of reliability are relevant to the current study. The PIMRS instrument has been used extensively as a tool for data collection in research, principal needs assessment, and principal evaluation. This discussion suggests that users should not evaluate the reliability of the instrument based on a single standard. Rather, users should select the instrument, the form of the instrument (i.e., self-report, teacher, or supervisor version), and the scale level based on the intended use of the data.

Results

This study was designed to assess the reliability of the Principal and Teacher Forms of the PIMRS under a variety of different conditions. We employed meta-analysis with a combination of original and extracted secondary data contained in 52 data sets obtained from 43 independent studies. We organize our presentation of the results first for the Principal Form and then the Teacher Form of the PIMRS.

Reliability of the PIMRS Principal Form

The results of our efforts to synthesize the reliability results for the PIMRS Principal Form are presented in Table 3. The combined sample used for the meta-analysis consisted of 2,508 principals (see Table 3). The whole scale alpha reliability estimate was .96. Reliability estimates for the three dimensions were .88 for Defines the School Mission, .91 for Manages the Instructional Program, and .93 for Develops a Positive School Learning Climate.¹⁰ These all meet a sufficiently high standard of reliability for use as a component in systems of principal evaluation.

The summary coefficients, rho-hat, for the 10 instructional leadership functions were consistently and substantially lower than estimates for the full scale and three dimensions. Estimates ranged from a low of .74 on Creates Incentives for Teachers to a high of .85 on Frames the School's Goals. These results suggest that the function-level subscales meet a standard of reliability sufficient for use in research and principal needs assessment, but not in principal evaluation.

Numerous scholars and practitioners have noted that features of the school context may shape or moderate the leadership behavior of principals (e.g., Belchetz & Leithwood, 2007; Goldring, Huff, May, & Camburn, 2008; Hallinger, 2011a). Differences in structural complexity and size create different challenges for principals who exercise instructional leadership in primary and secondary schools (Cuban, 1988). Similarly, the cultural context of the

school shapes both the formal job description and normative expectations held by teachers and parents for the principal (Hallinger & Leithwood, 1996; Hallinger & Lee, in press).

We took advantage of the breadth of our data sources to analyze the reliability of the PIMRS instrument across different school levels and cultural contexts in which the scale had been used. This allowed us to gain insight into how the scale responds under different conditions. The detailed results of these meta-analyses are displayed in Table 4.

Although there are some minor variations, the pattern of reliability results did not vary significantly either across different school levels or the two specific cultural contexts included in this analysis (i.e., United States and East Asia). At the same time, however, the sample size and number of studies both vary widely for the different cells in Table 4.¹¹ Furthermore, we note that it is not possible to generalize findings from two societies to all of "East Asia." Therefore, the results presented for East Asia should be interpreted as preliminary or indicative rather than conclusive.

Reliability of the PIMRS Teacher Form

For the PIMRS Teacher Form, the results of the meta-analysis are organized in terms of the three different reliability tests. The findings shown in Table 5 are based on the synthesis of data sets comprised of 8,153 respondents with an average of 19.6 teachers per school. The 18 data sets that employed Cronbach's alpha were comprised of 6,465 teachers, with an average sample size of 22 teachers per school. The four data sets that employed Ebel's test included 1,984 teachers, with an average sample size of 22 teachers per school. The 11 data sets containing gen theory coefficients included 3,615 teachers, with an average sample size of 11 teachers per school.

Although we were unable to mathematically combine the results of the three different reliability tests, it was of interest to understand how the magnitude of reliability coefficients differed depending on the test that was used. In order to gain insight into this issue, we proceeded to apply the Cronbach and gen theory reliability tests to our raw data. The results indicated that the gen theory formula tended to yield slightly higher coefficients than the results obtained from the Cronbach tests when applied to the same data (not tabled). We suggest that this is because of the capacity of the gen theory test to derive the reliability from teachers grouped by school and from item level rather than averaged responses. We consider the gen theory test to be the most accurate approach to representing the reliability of the

Table 4. Meta-Analysis of Principal Reliability Estimates by Cultural Context and School Level.

Variable	Whole Scale	Create Mission	Manage Instruction	Develop Climate	Frame Goals	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Professional Development	Incentives for Learners	Protect Instructional Time	N Principals Studies
Culture															
United States	.96	.90	.93	.92	.86	.80	.80	.84	.82	.77	.75	.83	.80	.79	84
Asia	.96	.87	.90	.94	.79	.75	.78	.77	.84	.83	.67	.77	.80	.83	471
School level															
Primary and K-12	.95	.86	.90	.92	.81	.78	.74	.80	.78	.75	.73	.79	.76	.77	157
Middle and secondary	.97	.89	.92	.94	.88	.81	.85	.85	.86	.82	.74	.85	.84	.83	128
															11

Table 5. Meta-Analysis of Teacher Reliability Estimates.

Number	Author(s)	Nation	N (Teachers)	N (Schools)	Teacher (School) Level	School Level	Whole Create Scale	Mission Instruction	Manage Instruction	Develop Climate	Frame Instruction	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Professional Development	Incentives for Learners	Protect Instructional Time
Cronbach's reliability in extracted data sets																				
1	Aste	United States	248	6	41	Primary	—	—	—	—	0.89	—	0.90	0.90	0.90	—	—	0.86	—	—
2	Chappellear	United States	213	18	12	Secondary	—	—	—	—	—	—	—	—	0.88	—	—	—	—	—
3	Clark	United States	356	90	4	Primary	0.94	—	—	—	—	—	—	—	—	—	—	—	—	—
4	Coldhart	United States	128	14	9	Primary	0.98	—	—	—	0.93	0.90	0.85	0.92	0.85	0.86	0.86	0.91	0.92	0.84
5	DuPont	United States	126	1	126	Primary/secondary	0.97	—	—	—	—	—	—	—	—	—	—	—	—	—
6	Fulton	United States	174	24	7	Secondary	—	—	—	—	0.92	0.91	0.87	0.91	0.85	0.77	0.83	0.90	0.87	0.90
7	Gallon	United States	514	34	15	Primary/secondary	—	—	—	—	0.94	0.93	0.92	0.93	0.91	0.90	0.89	0.94	0.92	0.92
8	Geiselman	United States	979	180	5	Primary	0.97	0.90	0.92	0.94	—	—	—	—	—	—	—	—	—	—
9	Hallinger (1983)	United States	104	10	10	Primary	—	—	—	—	0.89	0.89	0.90	0.90	0.90	0.81	0.78	0.86	0.87	0.84
10	Hart	International	524	18	29	Primary/secondary	1.00	0.90	0.85	0.87	—	—	—	—	—	—	—	—	—	—
11	MacNeill	Canada	903	56	16	Primary/secondary	0.97	—	—	—	0.89	0.86	0.88	0.86	0.83	0.76	0.86	0.90	0.84	0.85
12	Minus	United States	298	62	5	Middle	—	—	—	—	0.72	0.80	0.81	0.73	0.85	0.70	0.57	0.79	0.74	0.77
13	Nogoy	United States	424	76	6	Secondary	0.98	—	—	—	0.93	0.90	0.86	0.91	0.89	0.85	0.86	0.91	0.90	0.88
14	O'Day	United States	140	19	7	Primary	0.98	—	—	—	0.89	0.85	0.86	0.88	0.85	0.81	0.83	0.81	0.86	0.76
15	Roudebush	United States	358	12	30	Primary	—	—	—	—	0.84	0.82	0.78	0.85	0.72	0.72	0.83	0.91	0.83	0.80
16	Sheppard	Canada	624	50	12	Primary/secondary	—	—	—	—	0.93	0.90	0.89	0.92	0.86	0.82	0.84	0.90	0.88	0.86
17	Wafir	Maldives	225	5	45	Primary	0.98	—	—	—	0.92	0.90	0.88	0.91	0.84	0.86	0.86	0.91	0.88	0.91
18	Waters	United States	127	7	18	Primary	—	0.90	0.95	0.91	0.85	0.91	0.93	0.87	0.96	0.89	0.79	0.73	0.97	0.95
	Summary statistics	United States	6,465	682	22	Rho hat	0.98	0.90	0.92	0.94	0.91	0.89	0.87	0.89	0.87	0.82	0.84	0.90	0.88	0.87
Ebel reliability in extracted data sets																				
19	Howe	United States	393	34	12	Primary	0.95	—	—	—	0.94	0.96	0.96	0.95	0.94	0.96	0.96	0.97	0.94	0.95
20	Jonas	United States	405	29	14	Secondary	—	—	—	—	0.86	0.82	0.90	0.71	0.81	0.79	0.79	0.80	0.82	0.70
21	Letour	United States	412	26	16	Primary	—	—	—	—	0.76	0.74	0.76	0.65	0.67	0.80	0.89	0.80	0.70	0.80
22	Taraseha	Thailand	774	10	77	Secondary	—	—	—	—	0.96	0.95	0.93	0.95	0.94	0.90	0.94	0.94	0.93	0.95
	Summary statistics	United States	1,984	99	30	Rho hat	0.95	—	—	—	0.91	0.93	0.93	0.91	0.91	0.89	0.93	0.94	0.90	0.91

(continued)

Table 5. (continued)

Number	Author(s)	Nation	N (Teachers)	N (Schools)	Teacher (School)	School Level	Whole Create Scale	Manage Instruction	Develop Frame Climate	Develop Frame Goals	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Professional Development	Incentives for Learners	Protect Instructional Time
Generalizability theory reliability in raw data sets																			
23	Adam	United States	128	9	14	Secondary	0.98	0.97	0.96	0.97	0.95	0.93	0.91	0.83	0.92	0.93	0.88	0.82	0.91
24	Carson	United States	95	77	1	Secondary	0.97	0.96	0.95	0.92	0.92	0.89	0.88	0.90	0.85	0.80	0.88	0.81	0.84
25	Dale	United States	177	36	5	Primary	1.00	0.99	1.00	1.00	0.99	0.99	0.99	0.99	0.97	0.99	0.97	0.99	0.98
26	Fancera	United States	580	100	6	Secondary	0.96	0.93	0.94	0.94	0.87	0.89	0.87	0.86	0.61	0.84	0.89	0.90	0.84
27	Golding et al., 2012	United States	1,610	58	28	Primary/ secondary	0.99	0.98	0.98	0.98	0.96	0.94	0.95	0.94	0.92	0.94	0.92	0.91	0.87
28	Greb	United States	107	31	3	Primary	0.93	0.90	0.83	0.85	0.88	0.82	0.51	0.83	0.48	0.74	0.82	0.12	0.52
29	Lyons	United States	176	15	12	Secondary	0.98	0.93	0.95	0.98	0.90	0.85	0.78	0.92	0.94	0.92	0.97	0.93	0.88
30	Ponuasamy	Malaysia	105	14	8	Primary	0.99	0.95	0.96	0.97	0.86	0.94	0.91	0.89	0.91	0.91	0.89	0.91	0.89
31	Shafaeu	Malaysia	201	10	20	Primary/ secondary	0.99	0.96	0.94	0.99	0.94	0.85	0.89	0.93	0.94	0.97	0.97	0.98	0.94
32	Shatzer	United States	280	37	8	Primary	0.99	0.96	0.97	0.98	0.95	0.93	0.88	0.96	0.92	0.89	0.93	0.96	0.96
33	Wang	China	156	23	7	Secondary	0.99	0.97	0.98	0.99	0.93	0.95	0.96	0.95	0.94	0.96	0.94	0.96	0.95
Summary statistics			3,615	410	9	Rho hat	0.99	0.97	0.98	0.98	0.95	0.93	0.95	0.95	0.90	0.94	0.93	0.94	0.92

data. Moreover, these results provide a useful benchmark for interpreting the reliability results from the other tests displayed in Tables 5 and 6.

Our meta-analysis based on results of the gen theory test yielded a full-scale reliability of .99, with coefficients of .97 (Defines the School Mission), .98 (Manages Instruction), and .98 (Develops School Learning Climate) for the three dimensions (see Table 5). The combined reliability estimates for the 10 instructional leadership functions ranged from a low of .90 (Maintains High Visibility) to a high of .95 on several functions. Despite these high reliability coefficients, we noted considerable variability in the actual coefficients reported study by study. We further observed that these estimates are consistently higher than estimates for the Principal Form of the PIMRS, and as expected, the results obtained from Cronbach's test were slightly lower than from the gen theory test.

We followed the main analysis of the teacher data sets with analyses of reliability across school levels and cultural contexts.¹² These results bore similar patterns to the main analyses reported previously. The results of the gen theory test displayed in Table 6 indicate reliability levels consistently above .90 across both cultural contexts and all school levels. Again, the reliability estimates were somewhat lower for the 10 functions than the three dimensions and full scale, and results from the gen theory test were somewhat higher than those from Cronbach's test.

Discussion

This report sought to provide a comprehensive assessment of the reliability of the PIMRS, one of the most widely used survey instruments in the field of educational leadership and management (Hallinger & Heck, 1996; Leithwood & Jantzi, 2005; Robinson et al., 2008). Our research questions revolved around understanding the reliability of the PIMRS with respect to its two most commonly used forms (i.e., teacher and principal forms), its component subscales, and its use in different organizational and cultural contexts. In this final section, we interpret the results and then discuss limitations and implications of the study.

Interpretation of the Findings

Citations of instrument reliability included in published empirical reports are frequently based on a single validation study, or occasionally on several replication studies. In the current report, we were able to gather reliability data from 43 independent studies. These yielded 52 data sets for meta-analysis. The diversity of the studies allowed us to develop a comprehensive,

Table 6. Meta-Analysis of Teacher Reliability by Cultural Context and School Level.

Generalizability Theory Reliability: From Raw Data																			
Variable	Whole Scale	Create Mission	Manage Instruction	Develop Climate	Frame Goals	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Professional Development	Incentives for Learners	Protect Instructional Time	N Teachers	N Schools	N Teachers per School	N Studies	
Culture																			
United States	.99	.98	.98	.98	.96	.96	.94	.95	.95	.91	.95	.93	.93	.91	394	45	10	8	
Asia	.99	.96	.96	.99	.93	.94	.92	.92	.93	.94	.96	.95	.97	.94	154	16	15	3	
School level																			
Primary	.99	.98	.98	.99	.97	.97	.94	.97	.96	.93	.96	.96	.96	.94	339	29	11	6	
Secondary	.99	.95	.97	.98	.91	.93	.94	.93	.92	.91	.96	.92	.90	.91	263	40	13	6	
Cronbach's Alpha Reliability: From Extracted Data																			
Variable	Whole Scale	Create Mission	Manage Instruction	Develop Climate	Frame Goals	Communicate Goals	Supervise Instruction	Coordinate Curriculum	Monitor Progress	Highly Visible	Incentives for Teachers	Professional Development	Incentives for Learners	Protect Instructional Time	N Teachers	N Schools	N Teachers per School	N Studies	
Culture																			
United States	.97	.90	.92	.94	.90	.88	.87	.89	.88	.83	.83	.87	.91	.86	296	38	8	9	
Asia	.98	—	—	—	.90	.88	.85	.88	.87	.80	.81	.88	.86	.86	277	45	6	4	
School level																			
Primary	.97	.90	.92	.94	.91	.89	.87	.89	.87	.82	.84	.90	.88	.87	357	41	9	16	
Secondary	.98	—	—	—	.92	.90	.88	.91	.84	.86	.86	.91	.88	.91	225	5	45	1	

multidimensional view of the reliability of the PIMRS instrument and its use under different conditions.

Meta-analyses of reliability results were conducted separately for the Principal and Teacher Forms of the PIMRS. In each case, we provided analyses for the whole scale as well as its component subscales. The pattern of results was consistent with Gay's (1992) observation that even in highly reliable instruments, subscales tend to yield lower reliability than the full scale.

We concluded that the Principal Form of the PIMRS demonstrated moderately high to very high reliability, depending on the scale level being analyzed. More specifically, alpha coefficients were .96 for the whole scale, between .88 and .93 for the three dimension-level subscales, and between .74 and .85 for the 10 function-level subscales. Although the number of studies and sample sizes varied considerably, we found no substantial variation in the pattern of the results for the PIMRS Principal Form across school levels or between the United States and East Asia. Results of meta-analysis for the PIMRS Teacher Form demonstrated a consistently higher level of reliability for all three levels of scale measurement (i.e., $> .90$) as well as across the specified organizational and cultural contexts.

We earlier discussed standards for assessing instrument reliability and noted the importance of aligning the selected standard with the use of the data. Here we conclude that the PIMRS Principal Form produces data that meet or exceed the standard needed in both research and principal needs assessment (i.e., the primary purposes for which this form of the scale is employed). Principal self-report data obtained from the PIMRS could also be employed in principal evaluation, but we recommend using measurements obtained for the full scale and three instructional leadership dimensions rather than the 10 functions.

The results indicated that the PIMRS Teacher Form meets applicable standards of reliability required for use in both personnel assessment and research. This conclusion applies to all three levels of the scale and school levels. The pattern of results was particularly strong for the United States. However, we consider the results for East Asia tentative, due to the limited coverage of regional societies and small number of studies included in the sample.

Limitations and Implications

This report on the reliability of the PIMRS has relevance for both researchers and practitioners. While the data suggest that researchers should be confident in using the PIMRS for collecting data on principal instructional leadership, our analyses found slightly different patterns of results for the two forms of the

scales. In addition, the reliability estimates for the full scale and three dimensions were somewhat higher than for the 10 leadership functions. Users will wish to take note of these differences in order to determine their relevance *in relation to the purpose for which the data will be employed*. We emphasize this last point because our investigation revealed the importance of linking the standard of reliability to the use of the data. Thus, researchers and district policymakers or practitioners could draw different conclusions about which forms and scale levels of the PIMRS are best suited to their purposes.

Another implication emerged from our examination of the full body of literature in preparation for the analyses included in this report. We noted that in the early PIMRS studies researchers were generally assiduous in reporting the reliability estimates reported (i.e., Hallinger, 1983; Jones, 1987; Leitner, 1989; Taraseina, 1993) in the methods sections of their reports. Yet, surprisingly, only about one third of documented users included a reliability analysis in their own study procedures.

In our view, establishing reliability is especially important when an instrument is being used in a setting that differs in meaningful ways from the original validation site(s). For example, we earlier noted that the PIMRS has been employed in 26 different countries. Yet to date we were only able to obtain reliability estimates for data collected in 7 of those countries. We consider this an easily remedied, though potentially important oversight. Since most of these studies were master and doctoral dissertations, we recommend that supervisors be more stringent in making this a requirement in future graduate research studies (see also Hallinger, 2011b).

We also noted that several different tests have been employed by researchers to examine the reliability of the PIMRS. Although it was not an explicit goal of the study, we gained insight into both the complexity and criteria for testing the reliability of the two different forms of the PIMRS. Accordingly, we recommend that future users of the PIMRS Principal Form employ Cronbach's alpha and users of the Teacher Form use the gen theory formula included in this article.¹³ Although our comparison of results obtained from the Cronbach and the gen theory tests yielded relatively small differences, the latter represents a more accurate approach when employed with teacher data. This recommendation applies not only to testing the reliability of the PIMRS, but also other principal assessment scales as well. In our reading of the literature, we find that most researchers are using Cronbach's test of reliability under conditions that exceed its capacity to yield accurate results.

We limited the scope of this report to a study of the reliability of the PIMRS. Nonetheless, we acknowledge that high reliability is a necessary but insufficient feature of a valid instrument. Although we noted that several

studies have examined aspects of the PIMRS instrument's validity (Hallinger, 1983; Jones, 1987; O'Day, 1984; Taraseina, 1993; Wotany, 1999), an updated assessment of its validity is warranted. The authors are in the process of conducting a revalidation study (see Hallinger & Wang, 2013).

Moreover, as noted earlier, previous validity studies have focused exclusively on the PIMRS instrument's internal validity. In the future, researchers should also assess the external validity of the PIMRS. Tests could, for example, directly compare the results of the PIMRS with results obtained from other instruments and in relation to key organizational conditions (Kerlinger, 1966). Indeed, while examining the full set of PIMRS studies, we came across several that compared the PIMRS with instruments measuring other leadership constructs such as transformational leadership, emotional intelligence, cognitive styles, and leader authenticity (e.g., Dale, 2010; Greb, 2011; Meyer, 1990; Munroe, 2009; Sawyer, 1997; Shatzer, 2009; Tang, 1997). Although these were not explicitly framed as validation studies, findings from these studies could be employed to increase our understanding of the PIMRS's external validity.

We began this report with a quotation from Bridges (1982) who had implied that progress in our field would remain stunted in the absence of stronger conceptual frameworks (see also Bridges, 1967) and research instruments. We noted that the PIMRS was developed in direct response to the expressed need for research instruments that could contribute to a program of research on how leadership impacts learning. The current study was conducted in order to ascertain *empirically* whether 30 years after its development the PIMRS continues to warrant a place in the researcher's toolbox for addressing these high-priority lines of inquiry. Our findings clearly indicate that both forms of the PIMRS instrument meet high standards of reliability that are consistent with current usage. As a result, we believe that both researchers and practitioners should be able to make more informed choices concerning whether and how to use this instrument in assessing principal instructional leadership.

Acknowledgments

The authors also wish to acknowledge the invaluable assistance rendered by the scholars who contributed their raw data for the purposes of this study: Philip Adam, Craig Carson, Aaron Dale, Sam Fancera, Ellen Goldring, William Greb, Cheryl Long, Brendan Lyons, Myra Munroe, Wong Yau Nyau, Premavathy Ponnusamy, Ismail Shafeeu, Ryan Shatzer, Wang Sitong, and Tara Todd. The contributions of other researchers whose data we drew upon in this study are also appreciated and cited in the text. We also wish to acknowledge the helpful comments offered by Edwin M. Bridges and Ronald H. Heck to an earlier version of this manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors wish to acknowledge the funding support of the Research Grant Council (RGC) of Hong Kong for its support through the General Research Fund (GRF 841711).

Notes

1. This estimate is based on a combination of literature search and information obtained from the publisher of the scale. The instrument has been used by researchers from the United States, Canada, United Kingdom, Portugal, Mexico, Germany, Israel, Australia, China, Malaysia, Maldives, Vietnam, Philippines, Taiwan, Thailand, India, Pakistan, United Arab Emirates, Cameroon, Nigeria, Zambia, Oman, Bhutan, Turkey, New Zealand, and Saudi Arabia.
2. A new 22 item short form of the Teacher version of the PIMRS has also been developed. See Hallinger and Wang (2013) for details.
3. In addition, our current research will also yield a paper that updates and extends the validation of the Principal Instructional Management Rating Scale (PIMRS) for publication in 2013.
4. There were another 20+ studies that we were able to identify but unable to obtain. Typically these had been conducted in Asia where digital access to research documents was limited.
5. The discrepancy in the number of PIMRS studies referred to in the report (i.e., 177 vs. “over 200” mentioned earlier) reflects both the continuing completion of new studies and our discovery of new studies of which we had previously been unaware.
6. The publisher informed the authors that there is a current average of 10 PIMRS requests per month from researchers globally. It was noted that requests over the prior month had come from the United States, Malaysia, Nigeria, United Arab Emirates, Saudi Arabia, Pakistan, and Vietnam.
7. The discrepancy in totals presented is due to the fact that some researchers collected both teacher and principal response data. We treat these as separate data sets. Moreover, in some instances data were not usable.
8. In the Goldring et al. (2012) and Hallinger and Lee (in press) studies, the data were coded by school level and we were therefore able to analyze the reliability by school level in our subsequent research (i.e., see Table 5). In some other stud-

- ies (e.g., Wong, 2010), we lacked access to school codes, so we could not break the data down for school-level analyses.
9. We note that this procedure could not be applied to the secondary studies since the authors seldom reported the detailed level of statistical information required to compute reliability with this approach.
 10. Given the uniqueness (i.e., use of a short form of the instrument) and large size of the Hallinger and Lee (in press) data set described earlier in the methods section, we decided to check whether this data set might be distorting the overall results. Therefore, we ran the meta-analysis procedures for the principal data both with and without the Hallinger and Lee data set. There was no difference in the overall scale reliability (.96). When this unique data set was removed from the meta-analysis, each of the dimension-level rho-hat estimates increased by .01. We conclude therefore that the uniqueness of this particular data set did not distort the findings.
 11. The reader will note that the total number of studies identified in this study by school level may not match data provided in the previous tables. There were two reasons for this. The first reason was implied in note 7, which noted that certain studies identified as primary/secondary could be broken down by school level; some others could not. In addition, in order to increase the sample size of studies within the cells, we treated uncoded primary/secondary data sets as *primary* and middle schools as *secondary* in the analyses included in this table. Our rationale was that the primary schools far outnumbered the secondary schools in the mixed or uncoded studies (e.g., in the Hallinger, 1983, study there were eight primary schools, one middle school, and one high school). We treated middle schools as secondary schools due to the similarities in departmental structure.
 12. See note 7. The same rationale was applied to the categorization of studies for the analysis shown in Table 6.
 13. Researchers can contact the corresponding author at hallinger@gmail.com to obtain the SPSS syntax for this test.

References

- Adam, P. (2012). *The effect of principal instructional characteristics on the academic growth of lower socioeconomic students* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS.
- Anderson, J. (2006). *An analysis of the relationship of high school principals perceived instructional leadership management behaviors to school size and to student achievement* (Unpublished doctoral dissertation). Texas A&M University, Commerce, TX.

- Aste, M. (2009). *The relationship between principals instructional focus and academic achievement of high poverty students* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA.
- Babcock, C. (1991). *Perceived instructional leadership behavior: Gender differences, female principals, male principals* (Unpublished doctoral dissertation). Western Michigan University, Royal Oak, MI.
- Belchetz, D., & Leithwood, K. (2007). Successful leadership: Does context matter and if so, how? In C. Day & K. Leithwood (Eds.), *Successful principal leadership in times of change: An international perspective* (pp. 117-137). Dordrecht: Springer.
- Benoit, J. (1990). *Relationships between principal and teacher perceptions of principal instructional management and student achievement in selected Texas school districts with an emphasis on an index of effectiveness* (Unpublished doctoral dissertation). New Mexico State University, Las Cruces, NM.
- Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, 18(3), 34-64.
- Bridges, E. (1967). Instructional leadership: A concept re-examined. *Journal of Educational Administration*, 5(2), 136-147.
- Bridges, E. (1982). Research on the school administrator: The state-of-the-art, 1967-1980. *Educational Administration Quarterly*, 18(3), 12-33.
- Brown, J. (1991). *Leader behavior and school effectiveness* (Unpublished doctoral dissertation). University of Texas, Austin, TX.
- Carson, C. (2012). *The relationship between principal instructional leadership and student achievement in Missouri middle schools*. Unpublished manuscript, Saint Louis University, St. Louis, MO.
- Carr, T. (2011). *An examination of leadership styles in implementing instructional technology: A case study to examine the elementary school principal perspective* (Unpublished doctoral dissertation). North Central University, Minneapolis, MN.
- Chan, Y. C. (1993). *A study of principals' instructional leadership in Hong Kong secondary schools* (Unpublished master's thesis). Chinese University of Hong Kong, Hong Kong.
- Chappellear, T. C. (2011). *The relationship of teachers' perceptions of high school principal's monitoring student progress and student achievement in southeastern Ohio* (Unpublished doctoral dissertation). West Virginia University, Morgantown, WV.
- Chi, L. C. (1997). *Principal and teacher perceptions of principal instructional management behavior in Taiwan, Republic of China* (Unpublished doctoral dissertation). University of South Dakota, Bismarck, SD.
- Clark, I. (2009). *An analysis of the relationship between K-5 elementary school teachers' perceptions of principal instructional leadership and their science teaching efficacy* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis-St. Paul, MN.

- Coltharp, G. (1996). *The relationship between leadership behaviors and instructional leader effectiveness of building level administrators* (Unpublished doctoral dissertation). Kansas State University, Manhattan, KS.
- Condon, C., & Matthews, C. (2010). *Measuring principal performance: How rigorous are commonly used principal performance assessment instruments?* (A quality school leadership issue brief). Naperville, IL: Learning Point Associates.
- Corkill, R. (1994). *Instructional leadership behaviors perceived as important by teachers and principals for academic achievement in selected south Texas elementary schools* (Unpublished doctoral dissertation). Texas A&M University, College Station, TX.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York, NY: Wiley.
- Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools*. Albany, NY: SUNY Press.
- Dale, A. (2010). *A causal study examining how instructional leadership, transformational leadership, and the mediating effects of teacher self-efficacy influence the math achievement of third through fifth grade students as measured by the Maryland school assessment* (Unpublished doctoral dissertation). University of Maryland, Eastern Shore, MD.
- Dennis, C. (2009). *The relationship between principals' self-perceptions and teachers' perceptions of high school principals' instructional leadership behaviors in South Carolina schools* (Unpublished doctoral dissertation). University of South Carolina, Columbia, SC.
- Dunn, A. (2010). *A quantitative study of the perceptions of school leadership preparation in the State of Georgia* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.
- DuPont, J. P. (2009). *Teacher perceptions of the influence of principal instructional leadership on school culture: A case study of the American embassy school in New Delhi, India* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis-St. Paul, MN.
- Dwyer, D. (1986). Understanding the principal's contribution to instruction. *Peabody Journal of Education*, 63(1), 3-18.
- Ebel, R. (1951). Estimation of the reliability of ratings. *Psychometrika*, 16, 407-424.
- Fancera, S. F. (2009). *Instructional leadership influence on collective efficacy and school achievement* (Unpublished doctoral dissertation). Rutgers University, New Brunswick, NJ.
- Fraenkel, J., & Wallen, N. (1990). *How to design and evaluate research*. New York, NY: McGraw-Hill.

- Fulton, T. (2009). *High school principal instructional leadership behavior in high and low need and high and low achievement schools* (Unpublished doctoral dissertation). Dowling College, Oakdale, NY.
- Gallon, S. (1998). *A study of teachers' perceptions of the principal as an instructional leader: A comparative analysis of elementary, middle and senior high school principals* (Unpublished doctoral dissertation). Florida International University, Miami, FL.
- Gay, L. (1992). *Educational research*. New York, NY: MacMillan.
- Geiselman, S. (2004). *Predicting elementary school student achievement: The impact of principal gender and principal leadership skills* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- Gjelaj Merturi, E. (2010). *The perceptions of principal-based leadership practices on student reading achievement* (Unpublished doctoral dissertation). University of Phoenix, Phoenix, AZ.
- Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.
- Goldring, E., Cravens, X. C., Murphy, J., Porter, A., & Elliott, S. (2012, March). *The convergent and divergent validity of the Vanderbilt Assessment of Leadership in Education™ (VAL-ED): Instructional leadership and emotional intelligence*. Paper presented at the Annual Conference of Association for Education Finance and Policy, Boston, MA.
- Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: What influences principal practice? *Journal of Educational Administration*, 46(3), 332-352.
- Greb, W. (2011). *Principal leadership and student achievement: What is the effect of transformational leadership in conjunction with instructional leadership on student achievement?* (Unpublished doctoral dissertation). Marian University, Indianapolis, IN.
- Haack, M. (1991). *An investigation of the instructional management behaviors of principals in mid-sized public high schools in Iowa* (Unpublished doctoral dissertation). University of Northern Iowa, Cedar Falls, IA.
- Haasl, W. (1989). *Multiple perspectives of role definition of small school principals and their influence on effective instructional management* (Unpublished doctoral dissertation). University of Wisconsin, Madison, WI.
- Hair, J. F., Black, B., Babin, B., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Hakstian, A. R., & Whalen, T. E. (1976). A k -sample significance test for independent α coefficients. *Psychometrika*, 41(2), 219-231.
- Hallinger, P. (1983). *Assessing the instructional management behavior of principals* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Hallinger, P. (1990). *Principal instructional management rating scale*. Sarasota, FL: Leading Development Associates. (Original work published 1982)

- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329-351.
- Hallinger, P. (2011a). Leadership for learning: Lessons from 40 years of empirical research. *Journal of Educational Administration*, 49(2), 125-142.
- Hallinger, P. (2011b). A review of three decades of doctoral studies using the Principal Instructional Management Rating Scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly*, 47(2), 271-306.
- Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership and student achievement. *Elementary School Journal*, 96(5), 498-518.
- 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., & Lee, M.S. (in press). Exploring principal capacity to lead reform of teaching and learning quality in Thailand. *International Journal of Educational Development*.
- Hallinger, P., & Leithwood, K. (1996). Culture and educational administration. *Journal of Educational Administration*, 34(5), 4-11.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal*, 86(2), 217-248.
- Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, 94(3), pp. 328-355.
- Hallinger, P., & Wang, W.C. (2013). Measurement properties of the Principal Instructional Management Rating Scale. Technical Report, version 4.2. Hong Kong: Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education.
- Hart, J. (2006). *A cross-cultural study of principals' leadership behaviors according to teachers' perceptions* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA.
- Heck, R. H., Larson, T. J., & Marcoulides, G. A. (1990). Instructional leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly*, 26(2), 94-125.
- Henderson, N. (2007). *Teacher and principal perceptions of effective instructional leadership: An exploration of guiding practice and personal beliefs* (Unpublished doctoral dissertation). University of Texas, San Antonio, TX.
- Howe, W. (1995). *Instructional leadership in catholic elementary schools: An analysis of personal, organizational, and environmental correlates* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Jones, P. (1987). *The relationship between principal behavior and student achievement in Canadian secondary schools* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

- Kane, M. T., Gilmore, G. M., & Crooks, T. J. (1976). Student evaluations of teaching: The generalizability of class means. *Journal of Educational Measurement, 13*(3), 171-183.
- Kerlinger, F. (1966). *Foundations of behavioral research*. New York, NY: Holt Rinehart & Winston.
- Krug, F. (1986). *The relationship between the instructional management behavior of elementary school principals and student achievement* (Unpublished doctoral dissertation). University of San Francisco, San Francisco, CA.
- Lang, G., & Heiss, G. D. (1998). *A Practical guide to research methods*. New York, NY: University Press of America.
- Latham, G., & Wexley, K. (1981). *Increasing productivity through performance appraisal*. Menlo Park, CA: Addison Wesley.
- Lee, M. S., & Hallinger, P. (2012). Exploring the impact of national context on principals' time use: Economic development, societal culture, and educational system. *School Effectiveness and School Improvement, 23*, 461-482.
- Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. Nottingham, UK: National College of School Leadership.
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996 -2005. *Leadership and Policy in Schools, 4*(3), 177-199.
- Leithwood, K., & Montgomery, D. (1986). *Improving principal effectiveness: The principal profile*. Toronto: OISE Press.
- Leithwood, K. A., & Stager, M. (1989). Expertise in principals' problem solving. *Educational Administration Quarterly, 25*(2), 126-161.
- Leithwood, K., & Steinbach, R. (1991). Indicators of transformational leadership in everyday problem-solving of school administrators. *Journal of Personnel Evaluation in Education, 4*(3), 221-244.
- Leitner, D. (1989). *The effects of principal instructional management behavior on student learning: An organizational perspective* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- Long, C. L. (2008). *Instructional leadership: Perceptions of Mississippi career and technical education administrators and teachers* (Unpublished doctoral dissertation). Mississippi State University, Starkville, MS.
- Lyons, B. J. (2010). *Principal instructional leadership behavior, as perceived by teachers and principals, at New York State recognized and non-recognized middle schools* (Unpublished doctoral dissertation). Seton Hall University, South Orange, NJ.
- MacNeil, A. (1992). *Principal instructional management and its relation to teacher job satisfaction* (Unpublished doctoral dissertation). University of South Carolina, Columbia, SC.

- Mallory, B. (2002). *The relationship of teachers perceptions of instructional leadership and student achievement measured by state testing in ten North Carolina high schools* (Unpublished doctoral dissertation). East Carolina State University, Greenville, NC.
- Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370-397.
- Marshall, A. K. (2005). *Instructional leadership: Perceptions of middle school principals and teachers*. University of Texas, Pan American, Texas A&M University, Commerce, TX.
- Meek, J. (1999). *Relationship of principal instructional leadership and student achievement outcomes in North Carolina public elementary schools* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, NC.
- Merton, R. (1957). The role set: Problems in sociological theory. *British Journal of Sociology*, 8(2), 106-120.
- Meyer, T. (1990). *Perceived leader authenticity as an effective indicator of perceived instructional leadership behavior in middle level principals* (Unpublished doctoral dissertation). Andrews University, Berrien Springs, MI.
- Minus, E. (2010). *Leading in the middle: Leadership behaviors of middle level principals that promote student achievement* (Unpublished doctoral dissertation). George Washington University, Washington, DC.
- Munroe, M. D. (2009). *Correlation of emotional intelligence and instructional leadership behaviors* (Unpublished doctoral dissertation). University of Phoenix, Phoenix, AZ.
- Nogay, K. (1995). *The relationship of superordinate and subordinate gender to the perceptions of leadership behaviors of female secondary principals* (Unpublished doctoral dissertation). Youngstown State University, Youngstown, OH.
- Nunnally, J. (1978). *Psychometric methods*. New York, NY: McGraw-Hill.
- O'Day, K. (1984). *The relationship between principal and teacher perceptions of principal instructional management behavior and student achievement* (Unpublished doctoral dissertation). Northern Illinois University, Normal, IL.
- O'Donnell, R. (2002). *Middle-level principals' instructional leadership behaviors and student achievement* (Unpublished doctoral dissertation). Lehigh University, Lehigh, PA.
- Ogawa, R., & Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, 31(2), 224-243.
- Peariso, J. F. (2011). *A study of principals instructional leadership behaviors and beliefs of good pedagogical practice among effective California high schools serving socioeconomically disadvantaged and English learners* (Unpublished doctoral dissertation). Liberty University, Lynchburg, VA.

- Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration* (pp. 106-132). New York, NY: Longman.
- Ponnusamy, P. (2010). *The relationship of instructional leadership, teachers' organizational commitment and school effectiveness in Malaysia* (Unpublished master thesis). Universiti Sains Malaysia, Kuala Lumpur, Malaysia.
- Poovatanikul, V. (1993). *Analyses of the perceptions of Thai principals and teachers as to the principal's role as instructional leader* (Unpublished doctoral dissertation). Southern Illinois University, Carbondale, IL.
- Ratchaneeladdajit, R. (1997). *Perceptions of Thai principals and teachers toward the principals' role as instructional leaders in private schools in Bangkok, Thailand* (Unpublished doctoral dissertation). Southern Illinois University, Carbondale, IL.
- Reid, N. (1989). *Elementary school principals' behaviors in an urban school improvement program: A descriptive study of five principals* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.
- Robinson, V., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44(5), 635-674.
- Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods*, 11(3), 306-322.
- Roudebush, D. (1996). *Teacher's perceptions of building principals as instructional leaders: A comparative study in year-round multi-track and conventional school settings* (Unpublished doctoral dissertation). University of Denver, Denver, CO.
- Saavedra, A. (1987). *Instructional management behaviors of secondary administrators* (Unpublished master's thesis). Bukidnon State College, Malaybalay, Bukidnon, Philippines.
- Salvador, A. (1999). *Relationship between the teachers' and administrators' perceived school organizational health and administrators' instructional leadership in Our Lady of Perpetual Succor College, High School Dept., 1998-99* (Unpublished master's thesis). Ateneo de Manila University, Manila, Philippines.
- San Nicolas, E.Q. (2003). *Instructional leadership in Guam's public elementary schools* (Unpublished doctoral dissertation). University of San Diego, San Diego, CA.
- Sawyer, E. (1997). *The relationship between cognitive styles and instructional leadership* (Unpublished doctoral dissertation). University of Missouri, Columbia, MO.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350-353.

- Shafeeu, I. (2011). *A study of principals' instructional leadership in schools of Gaafu Dhaalu Atoll, Maldives* (Unpublished master's thesis). University of Malaya, Kuala Lumpur, Malaysia.
- Sheppard, B. (1996). Exploring the transformational nature of instructional leadership. *Alberta Journal of Educational Research*, 42(4), 325-344.
- Shatzer, R. H. (2009). *A comparison study between instructional and transformational leadership theories: Effects on student achievement and teacher job satisfaction* (Unpublished doctoral dissertation). Brigham Young University, Provo, UT.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149-155.
- Smith, S. (2007). *Principals' and teachers' perception of principals' instructional leadership* (Unpublished doctoral dissertation). University of South Carolina, Columbia, SC.
- Stevens, L. (1996). *Instructional leadership: A single district study of the multiple perceptions of central office administrators, principals, and elementary teachers* (Unpublished doctoral dissertation). State University of New York at Buffalo, Buffalo, NY.
- Tang, Tseng-Chung. (1997). *The relationships among transformational, transactional and instructional leadership behaviors of senior secondary school principals in Taiwan* (Unpublished doctoral dissertation). University of South Dakota, Bismarck, SD.
- Taraseina, P. (1993). *Assessing instructional leadership behavior of secondary school principals in Thailand* (Unpublished doctoral dissertation). Peabody College for Teachers of Vanderbilt University, Nashville, TN.
- Todd, T. (2006). *Instructional leadership in high schools: The effects of principals, assistant principals and department heads on student achievement* (Unpublished doctoral dissertation). Florida Atlantic University, Boca Raton, FL.
- van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement*, 1(3), 26-40.
- Villanova, R., Gauthier, W., Proctor, P., & Shoemaker, J. (1981). *The Connecticut school effectiveness questionnaire*. Hartford, CT: Bureau of School Improvement, Connecticut State Department of Education.
- Vinson, T. (1997). *The performance of secondary principals as instructional leaders in Mississippi public schools as perceived by superintendents, principals, and classroom teachers* (Unpublished doctoral dissertation). University of Mississippi, Oxford, MS.
- Wafir, N. (2011). *The role of the principal's leadership in contributing towards the professional development of teachers in schools* (Unpublished master's thesis). University of Nottingham, Nottingham, UK.

- Wang, S. (2011). *Leadership in Chinese junior secondary school: Discovering quantitative relationship between antecedent factors, instructional leadership, teacher perception and student academic achievement* (Unpublished master's thesis). Newcastle University, Newcastle Upon Tyne, UK.
- Waters, D. (2005). *Principal leadership behavior and 5th grade student achievement in high poverty schools* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA.
- Wong, Y. N. (2010). *Instructional leadership in Malaysia* (Unpublished master's thesis). Open University of Malaysia, Kuala Lumpur, Malaysia.
- Wotany, J. (1999). *The training, selection, and hiring practices of Cameroonian school leaders and the evaluation of the present effectiveness of their instructional leadership behaviors* (Unpublished doctoral dissertation). University of Missouri, Kansas City, MO.
- Yang, C. S. (1996). *Instructional leadership behaviors of elementary school principals in Taiwan, Republic of China* (Unpublished doctoral dissertation). University of Northern Colorado, Greeley, CO.
- Yogere, L. (1996). *Instructional leadership behaviors of principals as perceived by themselves and the teachers of Assumption, San Lorenzo, during the school year 1995-1996* (Unpublished master's thesis). Ateneo de Manila University, Manila, Philippines.

Author Biographies

Philip Hallinger is the Joseph Lau Chair Professor and Director of the Asia Pacific Centre for Leadership and Change. His areas of research interest include leadership, leadership development, and school change and improvement.

Wen-Chung Wang is a Chair Professor of Educational and Psychological Measurement, and the Director of Assessment Research Centre, The Hong Kong Institute of Education. His primary research interests include Rasch measurement, item response theory, and psychometrics.

Chia-Wen Chen is a Senior Research Assistant in the Assessment Research Centre. His research interests include item response theory, computerized adaptive test, and psychometrics. All three authors work at the Hong Kong Institute of Education.