International Journal of Educational Research 75 (2016) 115–133

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Developing a validated instructional leadership profile of Turkish primary school principals

Mehmet Sukru Bellibas[a,*], Okan Bulut[b], Philip Hallinger[c,e], Wen-Chung Wang[d]

[a] Department of Educational Sciences, Adiyaman University, College of Education, Room 70, Adiyaman, Turkey
[b] Department of Educational Psychology, University of Alberta, 6-110 Education Centre North, 11210 87 Ave. NW, Edmonton, AB T6G 2G5, Canada
[c] Department of Educational Administration and Policy, Chulalongkorn University, Faculty of Education, 254 Phayathai Rd., Bangkok 10330, Thailand
[d] Department of Psychological Studies, Hong Kong Institute of Education, 10 Lo Ping Rd., New Territories, Hong Kong
[e] Distinguished Visiting Professor, Department of Educational Leadership and Management, University of Johannesburg, Johannesburg, South Africa

ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form 29 July 2015
Accepted 11 October 2015
Available online xxx

1. Introduction

Contemporary research on school leadership has provided policymakers with increasingly persuasive evidence concerning the scope and means by which leadership contributes to student learning outcomes (Hallinger and Heck, 1996; Heck & Hallinger, 2015; Leithwood, Harris & Hopkins, 2008; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Robinson, 2006; Sebastian & Allensworth, 2012). This research increasingly points to instructional leadership as a critical role of principals who achieve promising results for school improvement (e.g., Leithwood et al., 2004, 2008; Robinson, 2006). Moreover, the impact of leadership appears to be most critical in schools that evidence the greatest need, schools that operate in challenging conditions (e.g., Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Day, 2009; Duke & Salmonowicz, 2010; Leithwood & Day, 2007; Leithwood, Harris et al., 2010; Murphy, 2008).

Within this body of research on the effects of school leadership on teaching and learning, the conceptualization of 'instructional leadership' developed by Hallinger and Murphy (1985) has become the most widely used model in empirical research (Hallinger, 2011; Leithwood et al., 2008; Robinson, 2006; Witziers, Bosker & Kruger, 2003). The research instrument associated with this conceptualization, the *Principal Instructional Management Rating Scale* (PIMRS; Hallinger, 1982, 1990), has been translated and used for research in more than 30 different countries (Hallinger, 2011; Hallinger & Wang, 2015). Nonetheless, Hallinger and Wang (2015) recently observed that very few 'international scholars' had undertaken systematic validation of the PIMRS for use in education contexts that differ significantly in system structure and social culture from the site of its original validation in the USA. In short, the validity of the constructs and items comprising the PIMRS should not be

* Corresponding author.
E-mail addresses: mbellibas@adiyaman.edu.tr (M.S. Bellibas), bulut@ualberta.ca (O. Bulut), hallinger@gmail.com (P. Hallinger), wcwang@ied.edu.hk (W.-C. Wang).

http://dx.doi.org/10.1016/j.ijer.2015.10.002

taken for granted in school systems where the principal's role is conceptualized differently, or where socio-cultural norms impose different expectations for leader behavior in schools (Fromm, Volante, Hallinger, & Wang, In press).

With this admonition in mind, the current study sought to develop and validate a Turkish language version of the PIMRS Teacher Form for use in research, policy and leadership practice. The researchers began with translation, content validation, and adaptation procedures designed to develop a Turkish version of the PIMRS Teacher Form. The resulting PIMRS Turkish Form was administered to a representative sample of 294 randomly selected teachers working in 23 randomly selected primary schools in six Turkish provinces. Construct validation tests were conducted at both the scale (e.g., confirmatory factor analysis, reliability) and item levels (e.g., Rasch analysis). We then drew upon the dataset gathered with the PIMRS Turkish Form to develop a 'preliminary national profile' of the instructional leadership of Turkey's primary school principals. This Turkish profile was then compared to an instructional leadership profile of principals from several other countries.

The significance of this study lies in two domains. First, policymakers throughout the world have become increasingly interested in developing reliable means of assessing the instructional leadership of school principals for related purposes of training and needs assessment, performance appraisal, and policy implementation evaluation (Hallinger and Wang, 2015). The current study demonstrates how a research tool designed in one country can be adapted and validated for use in another national context. Second, the study showed how research-based evidence on principal leadership could be compiled for local use as well as cross-national comparison. These dual contributions offer insights into ways that research tools can be adapted and validated for use in both basic and policy-oriented research in an increasingly diverse global education context.[1]

2. Theoretical perspective

In this section of the paper, we begin by presenting the conceptual framework developed by Hallinger and Murphy (1985) that guided the present investigation. Then we discuss the properties of the *Principal Instructional Management Rating Scale* (Hallinger, 1982,1990; Hallinger and Wang, 2015), the research tool used in this study.

2.1. Conceptual framework of instructional leadership

Traditionally scholars have conceptualized the school principalship in terms of three key roles: managerial, political, and instructional (e.g., Cuban, 1988). In the USA, as well as other countries, managerial behavior was generally viewed as predominant among these roles. Scholarship in educational leadership has, however, undergone a paradigm shift since 1980 when findings from research on 'effective schools' suggested that the instructional leadership role of the principal should be given greater emphasis (e.g., Bossert, Dwyer, Rowan, & Lee, 1982; Erickson, 1979; Hallinger & Heck, 1996; Hallinger & Murphy, 1985). This finding, though greeted with skepticism from some quarters (e.g., Cuban, 1988; Firestone & Herriot, 1982), subsequently stimulated the development of new leadership models aimed explicitly at improving the quality of teaching and learning in schools (e.g., Blasé, 1987; Bossert et al., 1982; Hallinger and Murphy, 1985; Leithwood et al., 2008; Marks and Printy, 2003; Robinson, 2006).

According to Leithwood et al. (2008), the most widely disseminated model of instructional leadership was developed by Hallinger and Murphy (1985). This conceptual framework proposed three dimensions of the principal's instructional leadership role. These included leadership practices aimed directly at improving teaching and learning processes (i.e., *Manages the Instructional Program*), as well as practices aimed at shaping the direction (i.e., *Defines the School's Mission*) and the climate (*Develops a Positive School Learning Climate*) of the school (Hallinger and Murphy, 1985). The model has remained relevant 30 years hence, as an evolving body of research on leadership for learning increasingly affirms effects of instructional leadership on student outcomes (e.g., Day, 2009; Hallinger, Bickman, & Davis, 1996; Hallinger and Heck, 1996; Heck, Larsen, & Marcoulides, 1990; Heck & Hallinger, 2010, 2014; Leithwood et al., 2008; Leithwood, Patten et al., 2010; Neumerski, 2013; Sebastian and Allensworth, 2012; Rigby, 2014; Robinson, 2006).

More specifically, this conceptual framework has been influential in guiding education policy and practice concerned with principal preparation, selection, and evaluation (Hallinger, 2011; Leithwood et al., 2008; Murphy, 2007). This model has also shaped the field's thinking about how leadership influences learning (Leithwood et al., 2008; Robinson, 2006) and the development of research tools used for studying and evaluating school leadership (e.g., Goldring et al., 2009; Hallinger & Murphy, 1985; Hallinger & Wang, 2015; Leithwood, Patten et al., 2010; Porter et al., 2010). Next, we examine the characteristics of the research instrument employed in this study, the PIMRS.

2.2. Principal instructional management rating scale

Drawing explicitly upon this conceptual model, Hallinger (1982), (1990) developed the Principal Instructional Management Rating Scale (PIMRS). The instrument follows the conceptual model outlined in Fig. 1 by incorporating three dimensions and 10 functions in a survey research instrument. These constructs are measured by 50 items, organized such that each function-level subscale is composed of five items (Hallinger & Murphy, 1985). The 50 PIMRS items are rated on a five-point Likert-type scale ranging from 1 (almost never) to 5 (almost always). The PIMRS has four forms, including a

[1] Hallinger and Lee, 2013 for example of how a national profile was developed in Thailand.

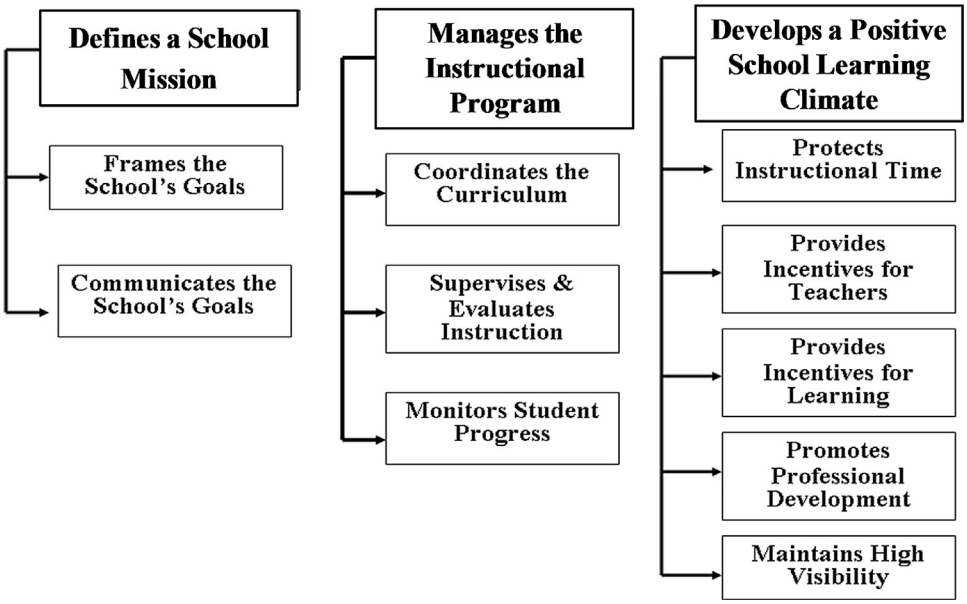

Fig. 1. PIMRS Instructional Leadership Model (Based on Hallinger & Murphy, 1985, p. 221).

principal self-assessment form, a teacher form, a teacher short form, and a supervisor form. Data collected through the PIMRS can be analyzed at the full scale, dimension or function levels (see Hallinger & Wang, 2015).

The PIMRS Principal and Teacher Forms have been subject to extensive testing for reliability and validity (Hallinger, 2011; Hallinger, Wang, & Chen, 2013; Hallinger & Murphy, 1985; Hallinger & Wang, 2015). A recent meta-analytic study of 40+ PIMRS studies found a full scale *alpha* reliability (Cronbach, 1951) coefficient of 0.96 for the PIMRS Principal Form and a Gen Theory reliability *rho hat* (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Kane, Gilmore, & Crooks, 1976) coefficient of 0.99 for the PIMRS Teacher Form (Hallinger & Wang, 2015). Dimension-level reliability coefficients exceeded 0.90 for all three dimensions on both forms. Function-level subscale reliability coefficients ranged from 0.75 to 0.86 for the PIMRS Principal Form and from 0.90 to 0.95 for the PIMRS Teacher Form (Hallinger et al., 2013; Hallinger & Wang, 2015).

Hallinger and Wang (2015) recently reported results that affirmed the internal validity of the PIMRS. Tests aimed at content validation supported the internal structure of the PIMRS. Internal validity of the composite items and subscales were further assessed using Rasch analysis. The results of Rasch analysis indicated that most of the items comprising the three dimensions of the PIMRS 'fit' the unidimensional assumption that was set as the criterion for assessing adequacy of subscale structure. In addition, application of differential item function (DIF) analysis across principals working at the primary and secondary levels found stability in empirical results obtained from the PIMRS (i.e., measurement invariance). Taken together, the results of validation testing indicated that the PIMRS meets commonly applied standards of reliability and internal validity (American Educational Research Association (AERA), 1999; Lang & Heiss, 1998; Lissitz and Samuelsen, 2007).

As the first well documented instrument aimed at measuring instructional leadership, use of the PIMRS has become ubiquitous among researchers around the world (Hallinger, 2011; Hallinger & Wang, 2015; Hallinger et al., 2013). Use of the PIMRS started in North America and has since expanded to include more than 375 studies conducted in 35 countries in Europe, Asia, South America, and Australia (Hallinger, 2011; Hallinger & Wang, 2015). These features of the PIMRS led leadership scholars in Turkey to become interested in adapting the scale for use in their national context.

2.3. Profiles of instructional leadership

For reasons discussed above, policymakers have been increasingly concerned with finding ways of assessing and developing system-wide instructional leadership capacity (see Fromm et al., in press; Hallinger & Lee, 2013, 2014; Lee & Hallinger, 2012). From the point of view of capacity building, system leaders have in recent years increased their investment in the preparation and professional development of school leaders. Indeed, all national leadership standards frameworks developed in the past decade around the world contain an explicit focus on instructional leadership (AITSL, 2014; Murphy, 2007; Walker and Ko, 2011). Moreover, the press to enhance instructional leadership capacity has also stimulated school systems to revamp recruitment and selection processes for principals and engage 'middle-level leaders' more explicitly in this domain (AITSL, 2014; Marks & Printy, 2003; Spillane, 2005; Walker and Kwan, 2012).

However, all of these efforts require validated measurement instruments capable of rendering accurate information concerning current leadership capacity, as well as the needs of leaders, and change in knowledge, skill and capacity over time (Catano & Stronge, 2006; Clifford & Ross, 2011; Goldring et al., 2009; Hallinger, 2012; Hallinger & Wang, 2015; Porter et al., 2010). Thus, instruments such as the PIMRS not only contribute to 'basic research' on school leadership processes, but also to policy related research and evaluation.

Hallinger and Lee (2013, 2014) recently employed the PIMRS in an effort to develop a national profile of the instructional leadership of Thailand's principals. This profile was used to describe patterns of instructional leadership practice among principals at different school levels and parts of the country. The results were employed both as a means of gaining insight into 'change in practice over time' as well as needs assessment with implications for both policy formation and professional preparation and development. Fromm and colleagues (In press) subsequently formulated and analyzed a national profile of principal instructional leadership in Chile. These efforts provided models for the analyses in the current study.

3. Method

This section begins by explaining the procedures used for translation of the PIMRS Teacher Form from English into Turkish language. Then we present the data collection and sampling procedures. This is followed by discussion of the tests used to evaluate the reliability and validity of the adapted form of the PIMRS. Finally, we discuss our approach to developing a national instructional leadership profile of principals in Turkish primary schools and comparing this profile to an international sample of primary school principals

3.1. Content validation of a Turkish form of the PIMRS

Prior to translation of the scale, we examined the suitability of the dimensions, functions and scale items for the Turkish context. We engaged a panel of scholarly and professional experts to examine the dimensions, functions, and items comprising the PIMRS. At the dimension level, the panel members agreed that all three dimensions (i.e., *Defines the School's Mission, Manages the Instructional Program, Develops a Positive School Learning Climate*) were relevant to the Turkish education context. However, one of the 10 function-level subscales, Coordinates the Curriculum, failed to achieve a consensus among panel members. Discussion among the panel members suggested that the Ministry of National Education's centralized control of curriculum content limited the relevance of this leadership function in the Turkish context. Thus, the panel recommended to exclude this function from the PIMRS-Turkish Form, pending a further review of the items.

The panel next examined each of the 50 items in the original PIMRS instrument. Although most of the items achieved a high degree of consensus on the criterion of relevance, several were highlighted as problematic. Specifically, panel members again questioned the suitability of items related to the principal's engagement with the curriculum. These were found in two of the 10 function-level subscales, *Coordinates Curriculum* (i.e., 5 items lacking relevance) and *Communicates the School's Goals* (i.e., 1 item lacking relevance).

Consequently, before translating the instrument, we removed six items that ask teachers to rate their principals on curriculum coordination, alignment and decision-making. In terms of conceptual structure, the PIMRS—Turkish form was comprised of three dimension-level constructs and nine function-level constructs. The total number of items was reduced from 50 to 44.

3.2. Translation procedures

We employed a 'back translation' method for transforming the original PIMRS into the Turkish language (Brislin, 1986). The 44 behavioral statements and the three demographic questions were translated from English to Turkish independently by three researchers fluent in both Turkish and English. Then, the Turkish copy was back translated to English and compared with the original form of the survey, in order to validate the translation process. The Turkish version was discussed with three professors in the Department of Educational Administration in a university in Turkey. Finally, the translated version of PIMRS was examined by six teachers. They were asked to read the survey and explain what they understood, what was clear and what was vague. Final revisions were made based on the feedback from the three professors and six teachers.

3.3. Data collection

The data for this study were collected from 294 primary school teachers in 23 schools across six provinces of Turkey. The provinces were selected to ensure a representative sample from different parts of the country, including the Eastern, Central and Western regions. This design was intended to maximize variation across types of schools across Turkey. The primary schools from each province were randomly selected.

Teachers were randomly selected from a list of the teaching staff in each school. Teachers with less than one year of experience in the school were eliminated. This was based on the assumption that they might not have had sufficient interactions with the principal to provide reliable information. In fact, the levels of experience of teachers who participated in the study ranged from three to 20 years.

Data collection was carried out by administering a paper and pencil form of the PIMRS to the recruited teachers in all schools. A researcher visited each school in the four provinces and distributed survey forms to the teachers. This gave a chance to further explain the purpose and scope of the study and provide answers to questions of participants. The completed documents were collected within seven days after distribution.

3.4. Data analysis

Data analysis procedures followed the three main goals of the study: to validate a Turkish form of the PIMRS, develop a national profile of the instructional leadership of Turkey's primary school principals, and compare the Turkish sample to principals in other national contexts. Our analytical strategy began by assessing the reliability of the data collected with the PIMRS Turkish Form. Next we proceeded with several tests designed to assess the internal validity of the PIMRS Turkish Form. We began by examining the 'fit' of data collected using the PIMRS Turkish Form against the item structure embedded in the original form of the PIMRS. If data collected with the PIMRS Turkish Form demonstrated a good fit to the PIMRS constructs, then we can conclude that the two forms of the PIMRS are measuring the same latent traits. Once comparability of constructs was established we developed a transformation table designed to align scores obtained by the 44 items comprising the PIMRS-Turkish Form with scores obtained by the 50 items comprising the original PIMRS.

Finally, we analyzed the instructional leadership profile of the Turkish principals in a two step process. First we conducted a descriptive analysis of the principals' instructional leadership across the PIMRS dimensions and functions. This revealed the relative degree of engagement of the principals in different dimensions of the instructional leadership role. Following this analysis, we compared the profile of the Turkish principals to an international comparison group in order to gain perspective on their pattern of instructional leadership practice.

3.4.1. Reliability analysis

We used a reliability formula based on Generalizability Theory (Cronbach et al., 1972; Kane et al., 1976) to test the reliability of the PIMRS Turkish Form. In an earlier study Hallinger et al. (2013) argued that this test afforded a stronger approach than Cronbach's *alpha* to testing the reliability of the PIMRS Teacher Form. They noted two specific advantages of the 'Gen Theory' test of reliability. First, it takes into account the hierarchical, nested structure of teacher data. Second, it utilizes the variability of item-level scores when calculating scale reliability (see Hallinger and Wang, 2015 for a detailed discussion of the formula).

3.4.2. Construct validation

To examine the functioning of items, two types of item-level analyses were conducted. First, classical item analysis techniques were used to obtain the average item scores and item-total correlations. Finally, scale-level analyses were conducted.

Traditional approaches to item analysis were initially employed to assess characteristics of the items within the conceptual model. These included analyses of test reliability, item-total correlation, and item difficulty. The item-total correlation (i.e., the bi-variate correlation between the item scores and test scores) is often used to assess the discrimination of an item: the higher the correlation, the higher the discrimination power an item has. Item difficulty in the classical item analysis is essentially the average of raw item scores for each item.

Because all items in the PIMRS use the same rating scale, the Rasch rating scale model (RSM; Andrich, 1978) was chosen to guide our analytical procedures. Rasch analysis was used to detect aberrant respondents as well as to examine the relationship between the items and the conceptual structure (i.e., the three PIMRS dimensions). This was necessary in order to ensure that scores obtained by the two versions would have a similar 'meaning'. Hallinger and Wang (2015) reported using Rasch analysis to transform the categorical item responses of the PIMRS to a location along a construct map. The distribution of locations of items on the construct map represents the structure of the original PIMRS instrument. Ideally the location of items in the adapted instrument would demonstrate a reasonable degree of alignment to the original structure.

In the RSM, the probability of person n (e.g., a teacher) endorsing score x on item i is defined as a function of the person's latent trait (θ_n), an overall item difficulty of item i (δ_i), and a step parameter τ_k (relative to δ_i) for each score k:

$$P_{nix} = \frac{\exp \sum_{k=0}^{x} (\theta_n - \delta_i - \tau_k)}{\sum_{j=0}^{J} \exp \sum_{k=0}^{j} (\theta_n - \delta_i - \tau_k)} \quad (1)$$

where $J+1$ is the number of categories in the items ($J=4$ for a five-point rating scale) and $\theta_n - \delta_i - \tau_k = 0$ when $k=0$. Higher item difficulty indicates greater difficulty in achieving a higher score on the item. The step parameters describe the relative thresholds of individual item scores. It should be noted that this approach to examining item difficulty diverges from classical item analysis. In Rasch analysis item difficulties and person scores from the RMS are based a logit scale ranging between −5 and +5.

An advantage of Rasch analysis is that person and item quality can be assessed by infit and outfit Z and mean square error (MNSQ) statistics. The MNSQ indicates the extent to which persons or items in the dimension fit the model. If the responses of a person or an item fit the model's expectation well, the expected value of MNSQ is 1. Thus, the closer the empirical value of MNSQ is to 1, the better the fit, that is, the normal person or higher item quality.

Wright, Linacre, Gustafen, and Martin-Lof (1994) recommended that for a 'good fit' the MNSQ for rating scale items should be located in a range from 0.6 to 1.4 and for a person less than 2.0. In this study, we detected aberrant respondents using the criterion of MNSQ > 2.0 in the person fit test. To examine item quality, we labelled items with MNSQ > 1.4 as items with poor fit. Items with MNSQ < 0.6 were labeled as acceptable.

An optimal test design includes items that represent a full range of item difficulty. This means that the instrument is capable of assessing low, moderate and high performance levels. Therefore, the first strategy is to ensure that all levels of difficulty are maintained when the number of items is reduced.

Item difficulty can be identified by examining the pattern of actual scores among a sample of principals. A Wright Map (Wright & Master, 1982) displays the distribution of item difficulty in relation to the distribution of teachers' ratings of their respective principals along a vertical line from the highest difficulty at the top to lowest difficulty at the bottom. The distribution of teachers' ratings of their principals is shown along the left hand side of the line and the distribution of item difficulty on the right hand side. The mean item difficulty is located at the zero point on the vertical line.

Using the Wright Map, we can clearly identify the distribution of items along these two parameters. The map also profiles the number of items and teachers located on each level of each of the two parameters. This information provides insight into how the item distribution of the scale (e.g., each dimension) changes according to the inclusion of different 'sets' of items. The optimal result is achieved when the means (i.e., item and rater) and the distributions (i.e., variance and shape) are similar.

We supplemented these item-level analyses of validity with analyses of the test structure at the scale level. Here we used confirmatory factor analysis (CFA) to examine the underlying factor structure of the PIMRS. We assessed a model comprised of the three dimensions and nine functions comprising the PIMRS Turkish Form. We employed Mplus 6 (Muthén and Muthén, 2010) for this test with a mean- and variance-adjusted weighted least squares (WLSMV) estimator.

Several indices were used to evaluate the fit of the CFA model. These included CFI, TLI, RMSEA, and WRMR. For TLI and CFI fit indices, values greater than 0.90 represented an acceptable fit, and values greater than 0.95 represented a good fit (Browne & Cudeck, 1993; Hu & Bentler, 1999). For RMSEA, values smaller than 0.07 were an indicator of a good fit (Steiger, 2007). Muthén and Muthén (2010) further asserted that a WRMR value close to one indicates good model fit if the other fit indices also indicate an adequate or good model fit. This sequence of item and scale level analyses was employed to assess the internal validity of the PIMRS Turkish Form.

3.4.3. Transformation of scale scores

As noted above, content validation of the new PIMRS Turkish Form reduced its length from the 50 items included in the original PIMRS to 44 items. The researchers wished to maintain comparability of scoring between the new and original PIMRS Teacher Forms so that the instrument could be used in cross-national studies and so that Turkish policymakers could benchmark their results against those obtained in other countries. Therefore, it was necessary to develop a 'transformation table' through which scores could be converted and/or compared.

After validation procedures were completed, we turned our attention to developing the transformation table. Based on the location of items on construct map developed for the original PIMRS, we calculated Rasch scores corresponding to each raw score in Turkish items. These Rasch scores were used as the anchor scale and used to compute corresponding expected scores for the original 50-item version of the PIMRS.

The resulting table will show scores on the Turkish version and equivalent scores in the English version. For example, the dimension *Defines the School's* Mission has a subscale length of 9 items. The range of possible scores is from 9 (i.e., if all responses are '1', almost never) to 45 (i.e., if all responses are '5', almost always), with a total of 37 possible total scores for the dimension. The Rasch score corresponding to each of the possible scores were computed and used as the anchor. According to item response theory, once Rasch scores and all item parameters in the original PIMRS Form version are obtained, the probability of responses can be calculated and the expected value of the raw score in the original version test can be predicted. This expected raw score in English version is equivalent to the corresponding Rasch score and raw score in the PIMRS Turkish Form.

3.4.4. Development of a national instructional leadership profile

In this study, we also sought to create a national profile that described the modal instructional leadership practice of Turkey's primary school principals. The performance profile portrays the principals' instructional leadership at both the dimension and function levels.

This study used two methods to compute the subscale scores. First, item scores were averaged within each instructional leadership subscale. Second, the RSM was used to estimate the subscale scores. The step parameters obtained in the item-level analysis were used to estimate the participants' scores in the nine subscales of the adapted PIMRS. Descriptive statistics of scale scores were provided to show the variation among the subscales of the adapted PIMRS. These procedures were then combined to present a profile of the sample of Turkish primary school principals on both the three instructional leadership dimensions and nine functions.

In order to gain additional insight into this profile, we used Rasch analysis to compare the Turkish results with results obtained in a multi-country sample. The multi-country sample was comprised of PIMRS data collected from 4377 teachers rating 651 principals in the USA, Malaysia and China. The dataset, obtained from the publisher of the PIMRS, was comprised of data collected in 13 independent PIMRS studies conducted between 2008 and 2012. The sample size of participating

Table 1
Generalizability theory-based test reliability results for the PIMRS Turkish Form.

Scale	*N*	Min	Max	Mean	SD	PIMRS Turkish Rel	Original PIMRS Rel[a]
Full scale	272	1	5	3.06	0.74	0.99	0.99
Dimension subscales							
Defines the school's mission	272	1	5	3.16	0.82	0.99	0.98
Manages the instructional program	278	1	4.6	2.95	0.81	0.99	0.98
Develops school learning climate	282	1.24	4.9	3.08	0.77	0.99	0.98
Function subscales							
Frames the school's goals	272	1	5	3.18	0.85	0.98	0.96
Communicates the school's goals	272	1	5	3.14	0.86	0.97	0.96
Supervises & evaluates instruction	278	1	4.8	2.99	0.80	0.96	0.94
Monitors student progress	278	1	4.8	2.90	0.94	0.98	0.95
Protects instructional time	282	1.2	5	3.35	0.76	0.95	0.95
Maintains high visibility	281	1	5	2.96	0.89	0.97	0.93
Provides incentives for teachers	281	1	5	2.83	0.92	0.98	0.91
Promotes teacher development	282	1	5	3.03	0.94	0.98	0.95
Provides incentives for learners	282	1	5	3.21	0.97	0.99	0.93

[a] Reported in Hallinger et al. (2013).

teachers in the 13 studies ranged from 95 to 1610. We employed Wright Maps to compare the pattern of teacher item responses in the Turkish and multi-country samples.

4. Results

This section is divided into three main sections. The first presents the results of the reliability analysis. This is followed by the results of the construct validation tests. Finally, we present the instructional leadership profile of Turkey's primary school principals.

4.1. Reliability of the PIMRS Turkish Form

The Generalizability Theory reliability and descriptive statistics for the Turkish data are shown in Table 1. The Generalizability Theory reliability coefficients for the 13 constructs ranged from 0.95 to 0.99. As indicated in the table, these were quite consistent with those obtained from the original version of the PIMRS (Hallinger and Wang, 2015). In sum, the PIMRS Turkish Form demonstrated a very high degree of reliability at all three construct levels (i.e., full scale, dimensions, functions). Indeed, the reliability results suggest that the PIMRS Turkish Form meets reliability standards suggested for multiple purposes including research, performance appraisal and needs assessment (Hallinger & Wang, 2015; Latham & Wexley, 1981; Nunnally & Bernstein, 1994).

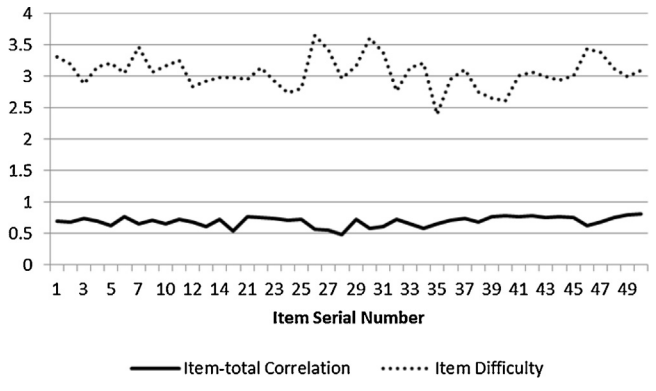

Fig. 2. Item-total correlation and item difficulty of the PIMRS-Turkish form.

Table 2
Item fit statistics for defines the school's mission (PIMRS Turkish Form).

Item label	*N*	Function label	Item difficulty	Standard error	Outfit MNSQ	Infit MNSQ	Goodness of fit
II_CSG_10	272	Comm goals	1.1	0.08	1.27	1.23	Acceptable
II_CSG_09	271	Comm goals	1.11	0.08	1.23	1.21	Acceptable
I_FSG_04	271	Frame goals	−0.65	0.09	1.04	1	Good
II_CSG_07	272	Comm goals	−0.62	0.09	0.87	0.83	Good
I_FSG_05	272	Frame goals	−0.23	0.08	0.75	0.73	Acceptable
I_FSG_01	272	Frame goals	−0.59	0.09	0.69	0.67	Acceptable
II_CSG_06	271	Comm goals	−0.16	0.08	0.68	0.64	Acceptable
I_FSG_03	271	Frame goals	0.49	0.08	0.66	0.64	Acceptable
I_FSG_02	272	Frame goals	−0.24	0.08	0.51	0.5	Acceptable

Table 3
Item fit statistics for manages the instructional program (PIMRS Turkish Form).

Item label	*N*	Function label	Item difficulty	Standard error	Outfit MNSQ	Infit MNSQ	Goodness of fit
III_SEI_15	277	Super & eval	−0.17	0.08	0.99	0.95	Good
V_MSP_23	278	Monitor prog	−0.43	0.08	0.95	0.95	Good
V_MSP_24	278	Monitor prog	−0.06	0.07	0.95	0.94	Good
III_SEI_14	277	Super & eval	−0.51	0.08	0.9	0.89	Good
V_MSP_21	277	Monitor prog	0.76	0.08	0.89	0.9	Good
III_SEI_13	278	Super & eval	0.33	0.07	0.8	0.8	Good
III_SEI_11	277	Super & eval	−0.49	0.08	0.77	0.72	Acceptable
V_MSP_25	276	Monitor prog	0.73	0.08	0.71	0.71	Acceptable
III_SEI_12	278	Super & eval	0.42	0.07	0.7	0.69	Acceptable
V_MSP_22	278	Monitor prog	0.03	0.07	0.64	0.63	Acceptable

Table 4
Item fit statistics for develops a positive school learning climate (PIMRS Turkish Form).

Item label	*N*	Function label	Item difficulty	Standard error	Outfit MNSQ	Infit MNSQ	Goodness of fit
VII_MHV_34	280	High visibility	0.82	0.07	1.94	1.81	Poor
VI_PIT_29	281	Protect time	−1.06	0.07	1.78	1.54	Poor
VI_PIT_28	282	Protect time	0.42	0.06	1.29	1.17	Acceptable
VI_PIT_27	282	Protect time	0.07	0.06	1.2	1.15	Good
VII_MHV_35	281	High visibility	1.21	0.07	1.03	1.06	Good
X_PIFL_46	282	Inc learning	−0.52	0.07	1	0.95	Good
VII_MHV_33	281	High visibility	−0.56	0.07	0.99	1.01	Good
VIII_PIFT_38	280	Inc teachers	0.83	0.07	0.99	0.87	Good
IX_PPD_43	281	Prof develop	−0.56	0.07	0.94	0.91	Good
IX_PPD_41	282	Prof develop	−0.6	0.07	0.92	0.89	Good
VI_PIT_30	281	Protect time	−0.23	0.06	0.87	0.88	Good
VI_PIT_26	282	Protect time	−0.41	0.07	0.84	0.88	Good
VII_MHV_31	281	High visibility	−0.25	0.07	0.83	0.85	Good
X_PIFL_47	282	Inc learning	−0.21	0.06	0.78	0.71	Acceptable
IX_PPD_42	282	Prof develop	−0.54	0.07	0.77	0.75	Acceptable
X_PIFL_48	282	Inc learning	0.4	0.06	0.73	0.73	Acceptable
IX_PPD_44	282	Prof develop	−0.24	0.06	0.67	0.65	Acceptable
VII_MHV_32	281	High visibility	0.38	0.06	0.63	0.64	Acceptable
VIII_PIFT_37	280	Inc teachers	−0.17	0.06	0.63	0.64	Acceptable
VIII_PIFT_36	280	Inc teachers	0.11	0.06	0.62	0.63	Acceptable
IX_PPD_45	281	Prof develop	−0.01	0.06	0.61	0.6	Acceptable
X_PIFL_49	282	Inc learning	0.46	0.06	0.6	0.59	Acceptable
X_PIFL_50	282	Inc learning	−0.36	0.07	0.6	0.6	Acceptable
VIII_PIFT_39	280	Inc teachers	0.47	0.06	0.52	0.51	Acceptable
VIII_PIFT_40	280	Inc teachers	0.56	0.06	0.51	0.5	Acceptable

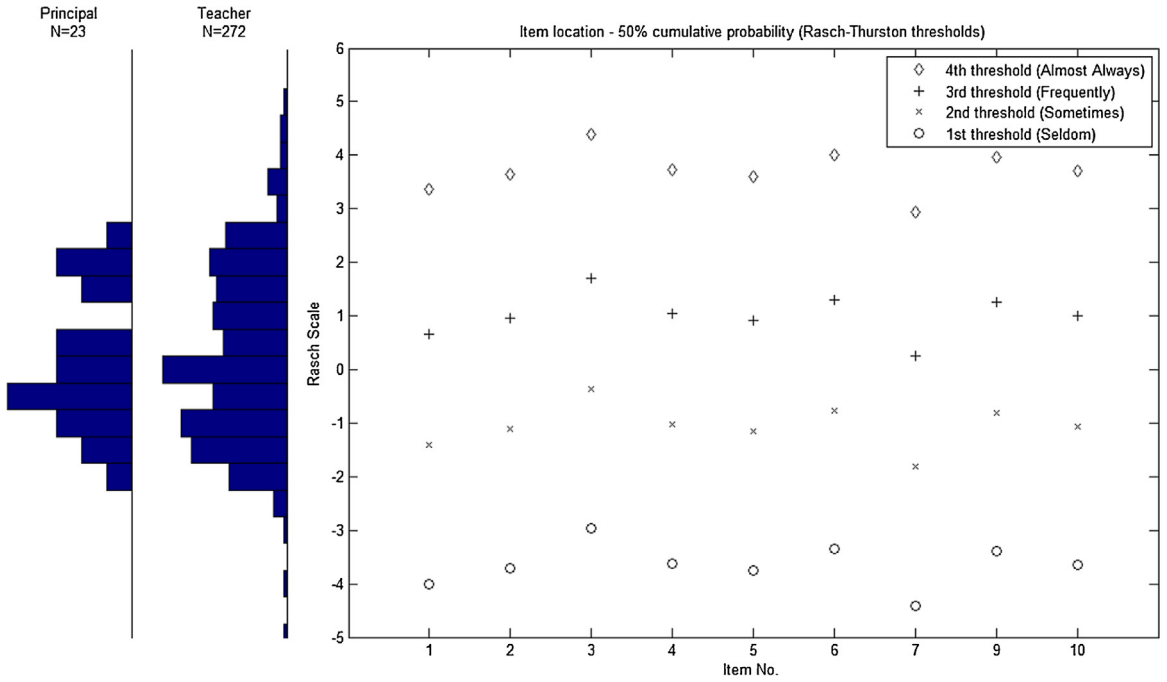

Fig. 3. Wright Map of persons and item thresholds on defines the school's mission (PIMRS Teacher Form).

4.2. Construct validation of the PIMRS Turkish Form

4.2.1. Item-level validation tests

The results for item-total correlation and item difficulty are shown in Fig. 2. All correlation coefficients are higher than 0.5 denoting that all items had sufficient discrimination to distinguish different teachers' evaluations based on the same framework.

We judged persons having an aberrant response when the outfit MNSQ for their responses exceeded 2.0. The results showed that 22 persons in *Defines the School's Mission*, 16 persons in *Manages the Instructional Program*, and 12 persons in *Develops a Positive School Learning Climate* were detected as aberrant respondents. After eliminating aberrant respondents for the different construct levels, we reran the reliability, item-total correlation, and item difficulty tests. All data reported in the figures and tables reflect the dataset without the aberrant respondents.

Tables 2–4 show item fit statistics for the three dimensions and their respective functions. In the first two dimensions (Tables 2 and 3), all MNSQ values were smaller than 1.4. This indicates that all item responses fit the structure of the original PIMRS version quite well.

For the *School Learning Climate* dimension (see Table 4), items 29 (VI_PIT_29) and 34 (VII_MHV_34), had a poor fit with an MNSQ of 1.94 and 1.78, respectively. In validation study, VII_MHV_34 also had a poor fit because principals appeared to practice these behaviors quite infrequently. Nevertheless, these items were retained in both versions of the PIMRS due to the conceptual proposition that instructional leadership impacts teaching and learning in both direct and indirect ways. Overall, item fit between the PIMRS Turkish Form and the original PIMRS was acceptable from a measurement perspective.

As noted above, our analysis of construct validity included a comparison of the theoretically derived 'construct map' and the empirically generated Wright Map for the three PIMRS dimensions (see Hallinger and Wang, 2015). When correspondence between the two maps is high, we have additional reason to believe that the instrument is measuring the construct as intended by the test developers. As indicated in Fig. 3, the Wright Map shows the distribution of persons on the left and the distribution of item thresholds on the left for the dimension, *Defines the School's Mission*. Higher scores along the vertical Rasch scale reflect greater item difficulty meaning that it was more difficult for principals to attain high scores on the particular item(s).

Comparison of the Wright Map (see Fig. 3) and the construct map for the first dimension, *Defines the School's Mission*, revealed a high degree of alignment. For example, the statement "Principal ensures that the importance of the school's goals is understood by discussing and reviewing them with staff" is considered 'Advanced' in the construct map. The fifth threshold (i.e., almost always) of this corresponding item in the Wright Map "Discuss the school's academic goals with

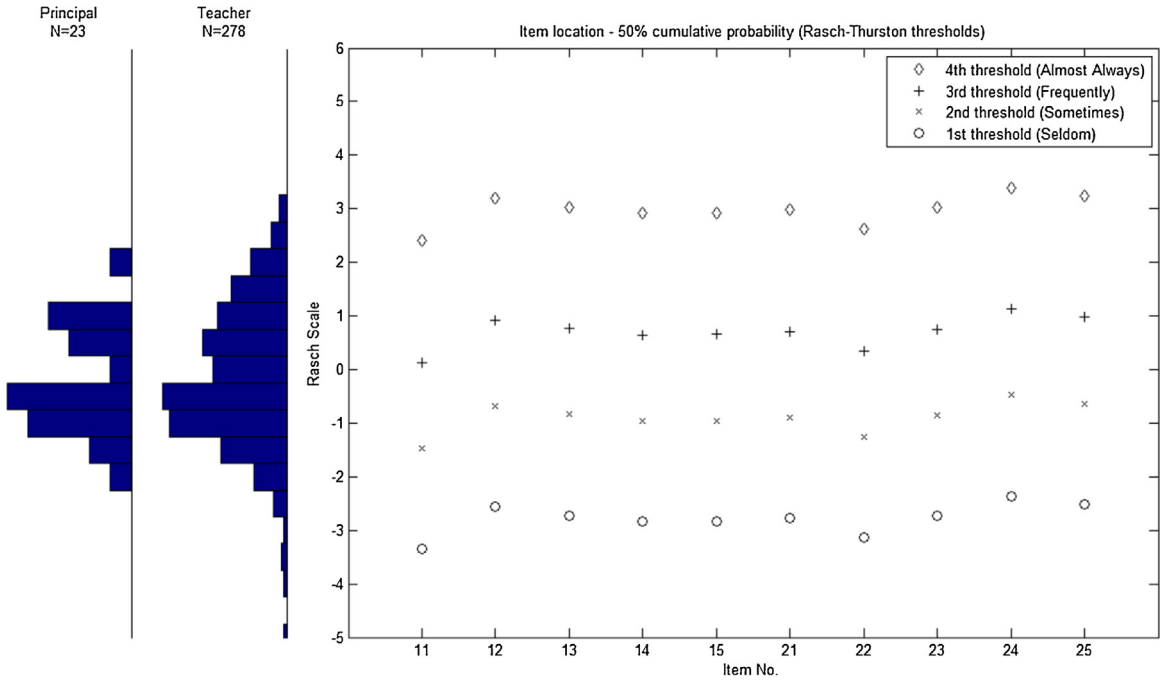

Fig. 4. Wright Map of persons and item thresholds on manages the instructional program (PIMRS Teacher Form).

teachers at faculty meetings" (4th threshold of item 7) is located in the 'Advanced' level, thereby suggesting alignment between their locations on the theoretical and empirical maps.

On the second dimension, *Manages the Instructional Program*, there was also evidence of agreement between the theoretical construct map and the empirical Wright Map (Fig. 4). For example, the statement "Principal provides instructional support to teachers and monitors classroom instruction through numerous informal classroom visits" is considered 'Advanced' on the construct map. A corresponding item (4th threshold of item 13), "Conduct informal observations in classrooms on a regular basis (informal observations are unscheduled, last at least 5 min, and may or may not involve written feedback or a formal conference)" is similarly located in the 'Advanced' level of the Wright Map

In the third dimension (*Develops a Positive School Learning Climate*), we found considerable agreement between the theoretical construct map and the empirical Wright Map (see Fig. 5), though there was some minor noise. This could be due to the fact that this dimension includes five functions as opposed to two and three for the first two dimensions. For example, the third threshold (Sometimes) of the item "Actively support the use in the classroom of skills acquired during in-service training" 2nd threshold of item 42) is considered 'Proficient' in the construct map, but is located in the 'Basic' level on the Wright Map. The second threshold (Seldom) of the item "Tutor students or provide direct instruction to classes" (1st threshold of item 35) is located in 'Proficient' level of the Wright Map but was proposed as 'Basic' in the construct map. Nonetheless, despite these discrepancies, the overall agreement between the construct map and the Wright Map on this dimension is still high enough to support construct validation for this dimension.

4.2.2. Scale-level validation tests

CFA results indicated that the nine-factor CFA model fit the PIMRS data very well. Both CFI (0.97) and TLI (0.97) indices were above the suggested cut-off value for a good model fit. The RMSEA and WRMR values for the model were 0.054 and 1.005, respectively. Although WRMR slightly exceeded the suggested cut-off value, it was still acceptable. Furthermore, all items had very high factor loadings within their subscales (see Table 5). Factor loadings ranged from 0.58 to 0.95, indicating that all of the PIMRS items were highly associated with their subscales.

Table 6 shows the correlations among the nine functions. The nine factors were moderately to highly correlated with each other. This is expected in an instrument where the sub-scales are conceptually related, and consistent with CFA results for the original form of the PIMRS (Hallinger and Wang, 2015). In conclusion, the nine-factor CFA model showed a good fit to the data gathered with the Turkish form of the PIMRS.

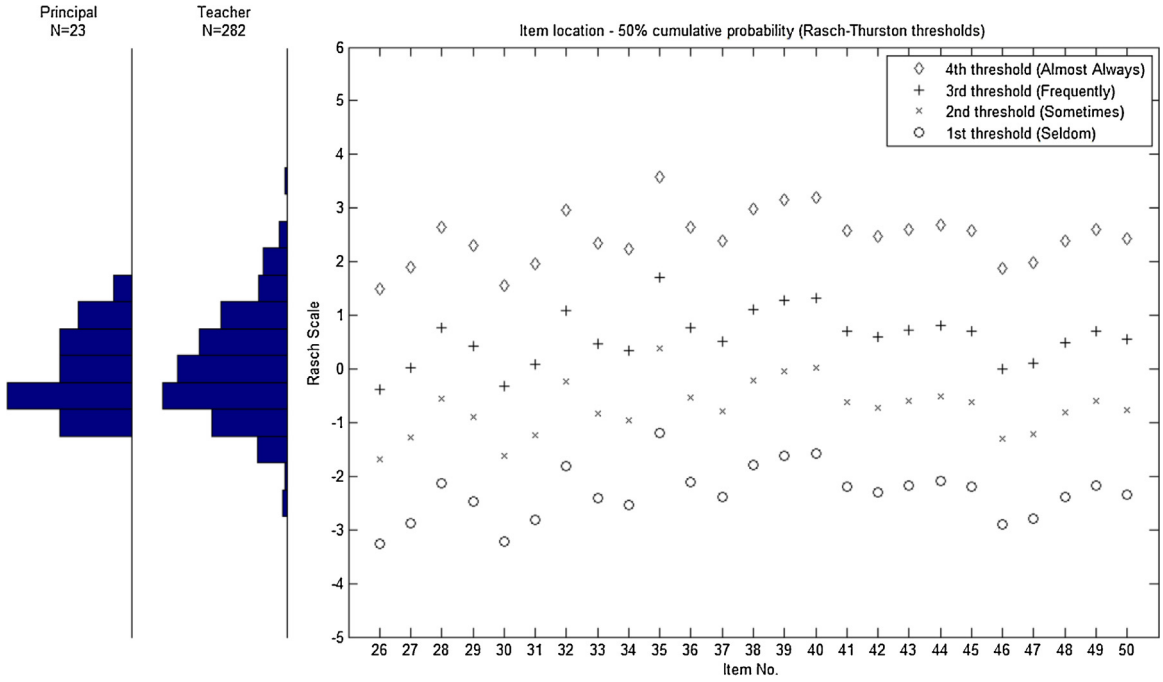

Fig. 5. Wright Map of persons and item thresholds on develops a positive school learning climate (PIMRS Teacher Form).

4.2.3. Developing a transformation table

As noted earlier, content validation resulted in a reduction in the number of items in the Turkish form of the PIMRS. This required us to construct a transformation table in order to equate scores obtained from the two forms of differing length. Drawing upon the item parameters of Rasch analysis in the English version of the PIMRS, a Rasch score for the corresponding Turkish and English items can be calculated.

The resulting transformation table[2] shows the scores obtained for a dimension or function for the Turkish version and equivalent scores for English version. For example, on the first dimension, *Defines the School's Mission,* a raw score of 31 on the Turkish version is equivalent to a raw score of 35 on the English version of the PIMRS. Using the transformation table, scores obtained in Turkish research can be directly compared to scores obtained in other studies that use the 50-item version of the PIMRS.

4.3. Developing an instructional leadership profile of Turkish primary school principals

Another purpose of the current paper was to develop a national profile with respect to principals' level of engagement in different functions and dimensions of instructional leadership. Fig. 6 displays mean values of each dimension and associated functions. According to the figure, principals appear most engaged in the *Defines the School's Mission* dimension. The mean values are higher than 3.0 (sometimes) for both functions in this dimension (i.e., *Frames the School's Goals*, *Communicates the School's Goals*).

Principals' engagement on the dimension *Develops a Positive School Learning Climate* demonstrated a mean value close to 3.0 (sometimes). Among the functions subsumed in this dimension, principals showed strongest engagement in *Protects Instructional Time* and most limited engagement in *Provides Incentives for Teaching.*

According to data presented in Fig. 6, teachers rated their principals least active in the dimension, *Manages the Instructional Program*. The mean value for this dimension was less than 3.0 (i.e., sometimes). According to the teachers'

[2] Due to space constraints the transformation table was omitted from this paper. However, it can be obtained from the publisher of the PIMRS at hallinger@gmail.com.

Table 5
Standardized factor loadings for the nine PIMRS (Turkish) subscales.

Subscales		Item	Item Description	λ
S1	Frames the school's goals	1	Annual school-wide goals	0.86
		2	Framing goals in terms of staff	0.85
		3	Needs assessment	0.89
		4	Use data on student performance	0.84
		5	Develop goals	0.77
S2	Communicates the school's goals	6	Communicate the school's mission	0.89
		7	Discuss goals with teachers	0.76
		8	Academic goals are reflected	0.82
		9	Refer to the school's goals or mission	0.76
S3	Supervises and evaluates instruction	10	Ensure priorities of teachers	0.83
		11	Review student work	0.76
		12	Conduct informal observations	0.68
		13	Point out specific strengths	0.81
		14	Point out specific weaknesses	0.59
S4	Monitors student progress	15	Meet individually with teachers	0.88
		16	Discuss academic performance	0.86
		17	Assess progress towards goals	0.85
		18	Inform teachers	0.84
		19	Inform students	0.85
S5	Protects instructional time	20	Limit interruptions	0.69
		21	Ensure that students not being called	0.67
		22	Missing instructional time	0.58
		23	Encourage teachers	0.88
		24	Limit the intrusion	0.72
S6	Maintains high visibility	25	Talk informally with students	0.68
		26	Visiting classrooms	0.84
		27	Attend in extra activities in school	0.75
		28	Cover late or substitute teachers	0.66
		29	Tutor students	0.76
S7	Provides incentives for teachers	30	Reinforce superior performance	0.81
		31	Compliment teachers privately	0.86
		32	Acknowledge teachers' performance	0.81
		33	Reward special efforts by teachers	0.93
		34	Create professional opportunities	0.93
S8	Promotes professional development	35	In-service activities for school goals	0.90
		36	Support the use of classroom skills	0.91
		37	Obtain staff participation	0.88
		38	Lead teacher in service activities	0.90
		39	Set aside time at faculty meetings	0.87
S9	Provides incentives for learning	40	Recognize superior students	0.79
		41	Honor academic accomplishments	0.86
		42	Recognize superior achievement	0.92
		43	Contact parents	0.95
		44	Support teachers actively	0.95

perceptions, on average, the Turkish principals do not appear strongly engaged on the two functions, *Supervises and Evaluates Instruction* and *Monitors Student Progress*.

This 'preliminary profile' provides a picture of relative 'levels of engagement' of Turkish primary school principals in different dimensions of the instructional leadership role. Drawing upon this profile, system leaders could target different needs for capacity development or policy analysis. For example, the pattern of low engagement on the dimension *Manages the Instructional Program* could imply a need for additional training or system support. While development of a national profile of principal instructional leadership has potential utility within Turkey, it is also of interest to understand how this profile compares to profiles developed in other countries.

Therefore, in the next step in this research, we sought to compare this 'Turkish profile' with similar data gathered in several other national contexts (see Hallinger and Wang, 2015). We generated Wright Maps comparing the item distributions

Table 6
Correlations among the nine subscales of PIMRS obtained from the CFA model.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	1								
S2	0.66	1							
S3	0.60	0.61	1						
S4	0.53	0.64	0.58	1					
S5	0.42	0.42	0.44	0.43	1				
S6	0.39	0.43	0.47	0.47	0.35	1			
S7	0.45	0.50	0.53	0.55	0.40	0.46	1		
S8	0.56	0.61	0.58	0.57	0.46	0.47	0.57	1	
S9	0.42	0.53	0.45	0.51	0.35	0.42	0.48	0.54	1

Note: S1: framing the school's goals; S2: communicating the school's goals; S3: supervision and evaluation of instruction; S4: monitoring student progress; S5: protecting instructional time; S6: visibility; S7: incentives to improve teaching; S8: promoting instructional improvement and professional development; S9: providing incentives for learning.

on the three dimensions of the PIMRS for Turkey and the multi-country sample described in the method section of the paper. Table 7 shows the transformed Rasch scores of the Turkish and multi-country teacher samples.

The table shows that in comparison to other countries (i.e., China, Chile, Malaysia, Maldives and USA), teacher perceptions of the Turkish principals' level of engagement in instructional leadership is quite low. Indeed, the Turkish principals received the lowest ratings among the six societies on the total score, as well as the dimension- and function-level scores. This trend of

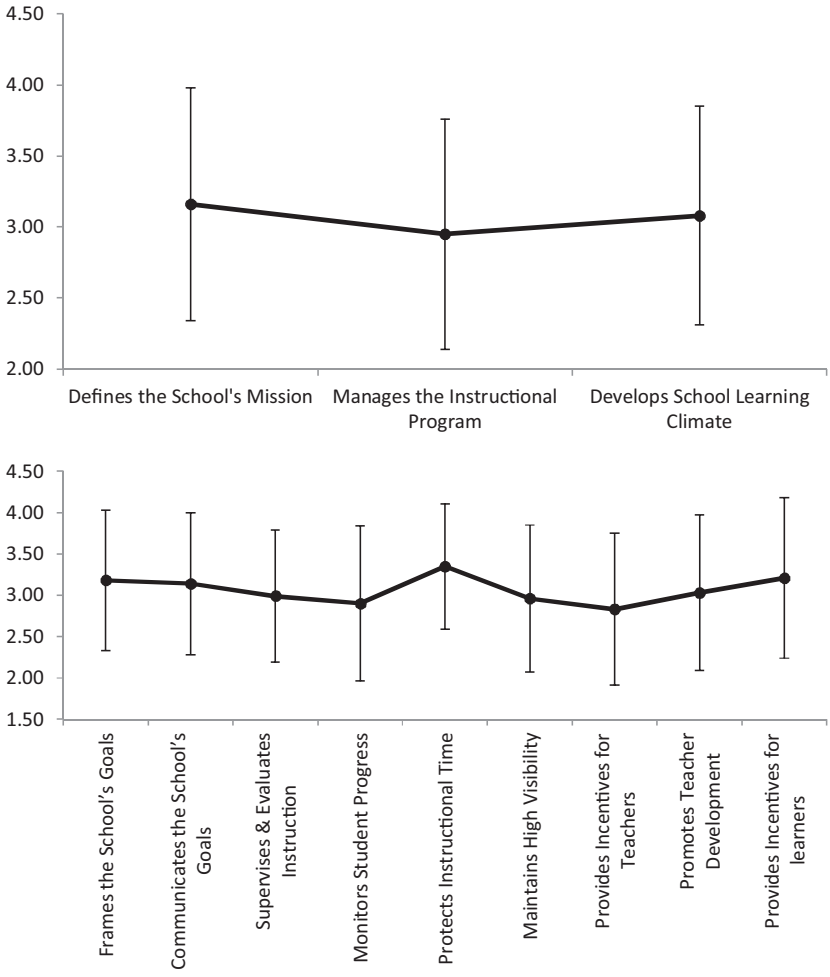

Fig. 6. Instructional leadership profile of Turkish primary school principals based on the PIMRS dimensions (above) and functions (below).

Table 7
Mean scores by total scale, each dimension, and each function in the different countries and the international area.

Name	Turkish	Chile	China	Malaysia	Maldives	USA	Composite international
Total scale	3.06	3.55	4.77	4.18	3.79	3.83	3.87
Dimension							
Defines the school mission	3.16	3.93	4.79	4.23	3.88	4.07	4.09
Manages the instructional program	2.95	3.59	4.81	4.21	3.87	3.87	3.91
Develops a school learning climate	3.08	3.43	4.75	4.14	3.70	3.61	3.67
Function							
Frames the school goals	3.18	4.07	4.82	4.20	3.95	4.14	4.15
Communicates the school goals	3.14	3.79	4.75	4.26	3.81	3.97	4.00
Supervises and evaluates instruction	2.99	3.55	4.85	4.26	3.96	3.88	3.93
Coordinates the curriculum	2.90	3.77	4.79	4.14	3.90	3.90	3.94
Monitors student progress	3.35	3.47	4.78	4.21	3.75	3.76	3.80
Protects instructional time	2.96	3.39	4.81	4.15	3.66	3.74	3.79
Maintains high visibility	2.83	3.41	4.57	4.20	3.78	3.29	3.37
Provides incentives for teachers	3.03	3.09	4.77	3.95	3.54	3.46	3.52
Promotes professional development	3.21	3.67	4.83	4.25	3.77	3.83	3.88
Provides incentives for learners	3.06	3.60	4.74	4.13	3.76	3.69	3.74

Note: sample size: Turkey = 294; Chile = 595; China = 156; Malaysia = 98; Maldives = 201; USA = 3915; international composite = 4377.

lower scores was most pronounced on the second dimension, *Manages the Instructional Program*, where the Turkish principals received significantly lower scores than the international comparison group (see Fig. 7).

This pattern of results suggests two possibilities. First it is possible that the Turkish teachers were simply more stringent in evaluating the instructional leadership of their principals. Alternatively, the scores may indicate that the Turkish principals were indeed less active in engaging this role than principals in the other six countries. We suggest that the consistency with which PIMRS scores for the Turkish principals lagged behind principals in all six of the comparison societies offers tentative support for the latter interpretation.

Figs. 8–10 show Wright Maps comparing the Turkish and multi-country teacher distributions for each of the three PIMRS dimensions. The Turkish and multi-country teacher distributions are shown in the first and second columns of the left-hand side of the Rasch logit scale, respectively. The item distribution is shown on the right-hand side. Since in our analysis the item structure was fixed to the structure of the original English version of the PIMRS, the item distribution on the Wright Maps was *not* re-calibrated to the Turkish version. Hallinger and Wang (2015) already demonstrated that the item distribution was well aligned with the theoretical structure of the instrument so there was no need to repeat that in this study. Instead, we focus on comparison between the distributions of the Turkish sample with the multi-country population.

On the first dimension, *Defines the School's Mission* (see Fig. 7), there was a high frequency bar located in the high score area. Except for this bar, other multi-country teachers had a normal distribution shape. Nonetheless, we note that the distribution of the multi-country population is somewhat higher than the Turkish profile, which was closer to the median. The Turkish teachers had a probability of around 0.5 to evaluate their principal with the median score. On the other hand, the distribution of Rasch scores in the Turkish sample was wide enough to suggest that an adequate range of performance levels on principal instructional leadership had been collected in this study.

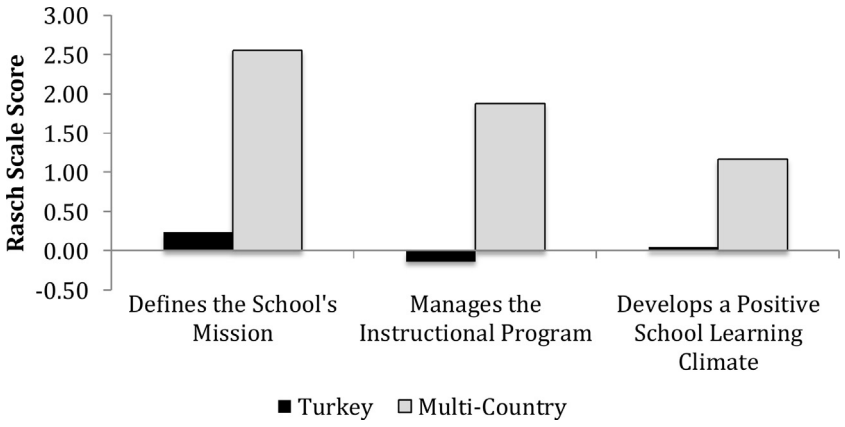

Fig. 7. Comparison of average Rasch scores from the three PIMRS dimensions between Turkey and multi-national data.

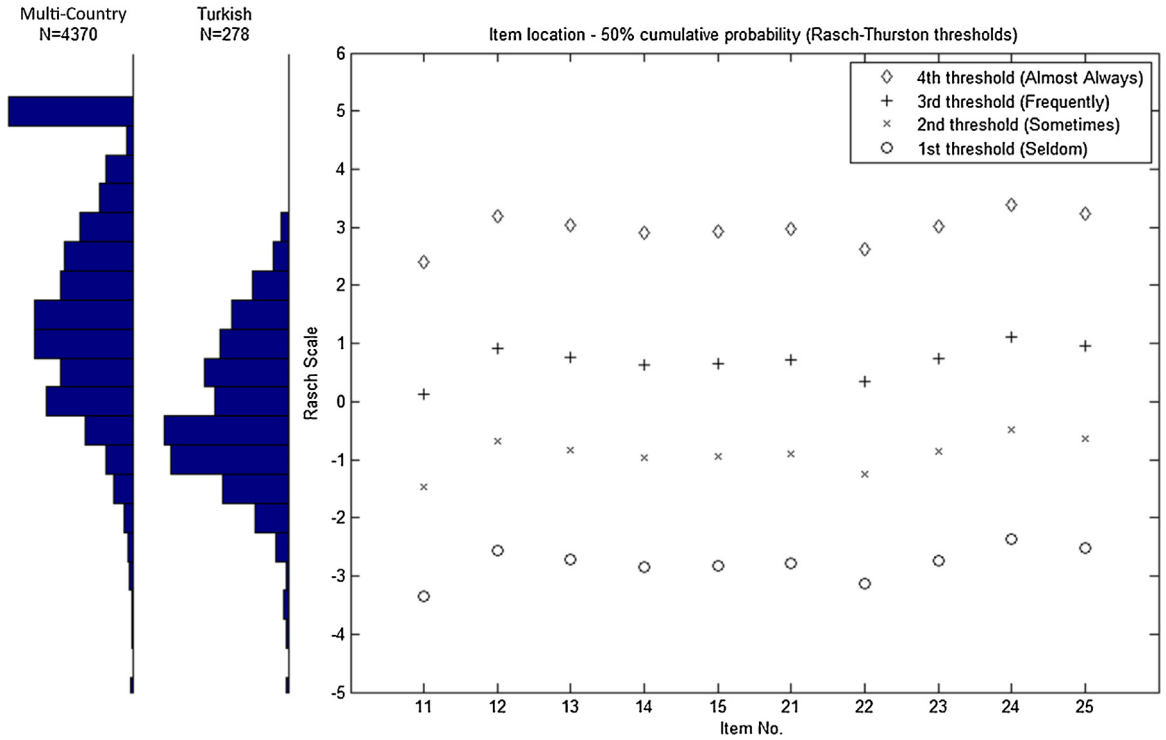

Fig. 8. Wright Map comparing Turkish and multi-country data for defines the school's mission (PIMRS Teacher Form).

Figs. 8 and 9 show Wright Maps for the dimensions *Manages the Instructional Program* and *Develops a Positive School Learning Climate*, respectively. Here we note a similar trend of higher evaluations in the multi-country sample, though at a lower frequency than for the first dimension. Again comparison of the distribution of Rasch scores on these dimensions suggests that the Turkish principals demonstrated a somewhat lower trend of scores across the PIMRS.

5. Discussion

Consistent with the global trend evidencing increased interest in strengthening school leadership (e.g., AITSL, 2014; Day, 2009; Leithwood et al., 2008), the current study was aimed towards developing the capacity to create a national instructional leadership profile of Turkish primary school principals. The first steps towards this goal entailed the translation and validation of the PIMRS (Hallinger, 1982, 1990), an appraisal instrument that has been employed widely in research and practice for more than 30 years. Given differences in socio-cultural and organizational context between the USA and Turkey, we deemed it necessary to validate the PIMRS for use in Turkey.

Our validation procedures sought to ensure that both the conceptual structure and items comprising the Turkish form of the PIMRS 'fit' the role instructional leadership expected of Turkish principals. After completing the validation tests, we created a preliminary profile of the instructional leadership of Turkish primary school principals and then compared this with an international sample of principals. In this final section of the article, discuss the findings and highlight implications for research and practice.

5.1. Summary and interpretation of findings

Content validation procedures resulted in the elimination of one subscale (Coordinates the Curriculum) and a total of six items from the original English language version of the PIMRS. The resulting PIMRS Turkish Form is, therefore, comprised of the same three instructional leadership dimensions, nine of the original 10 functions (see Fig. 1), and 44 of the original 50 items contained in the standard version of the PIMRS. A Gen Theory test of reliability yielded high reliability for the full scale (0.99), three dimensions (0.99) and 9 functions (i.e., ranging from 0.95 to 0.99). These reliability coefficients are equal to or even slightly higher than results reported in a recent meta-analytic study of the reliability of the original PIMRS instrument (Hallinger and Wang, 2015; Hallinger et al., 2013).

Both item-level and scale-level analyses were conducted to assess the construct (internal) validity of the PIMRS-Turkish Form. Rasch analysis confirmed that items in the PIMRS Turkish Form continues to provide adequate measurement of the composite constructs. CFA yielded high factor loadings indicating that all of the PIMRS items were highly associated with

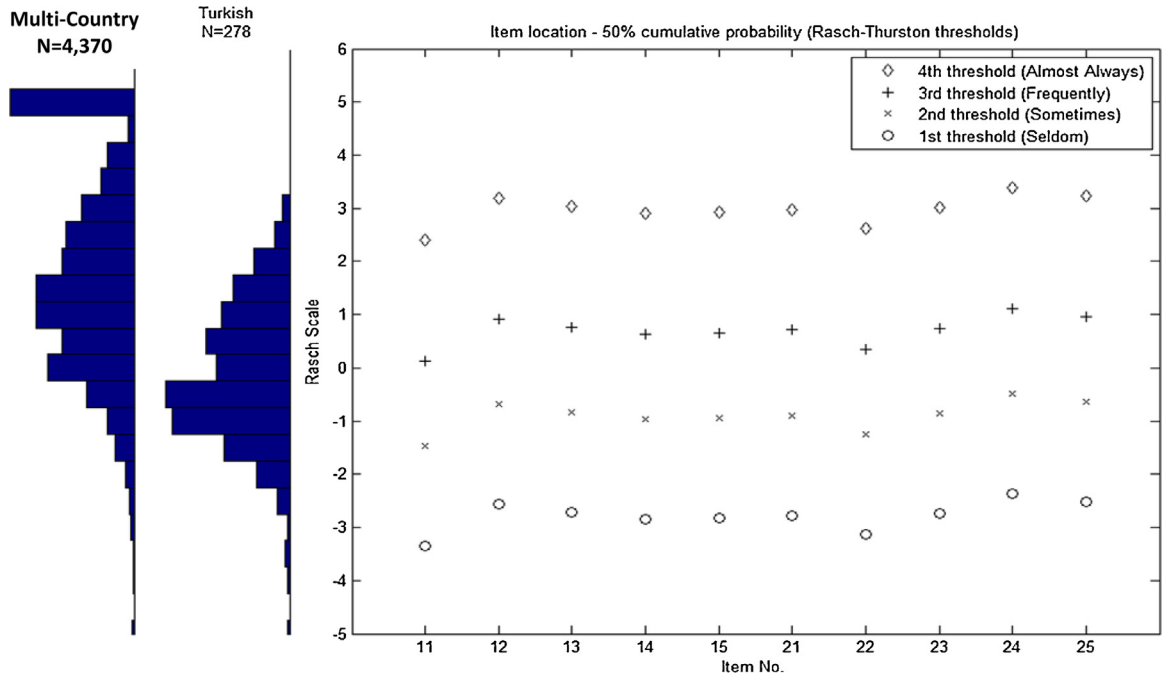

Fig. 9. Wright Map comparing Turkish and multi-country data for manages the instructional program (PIMRS Teacher Form).

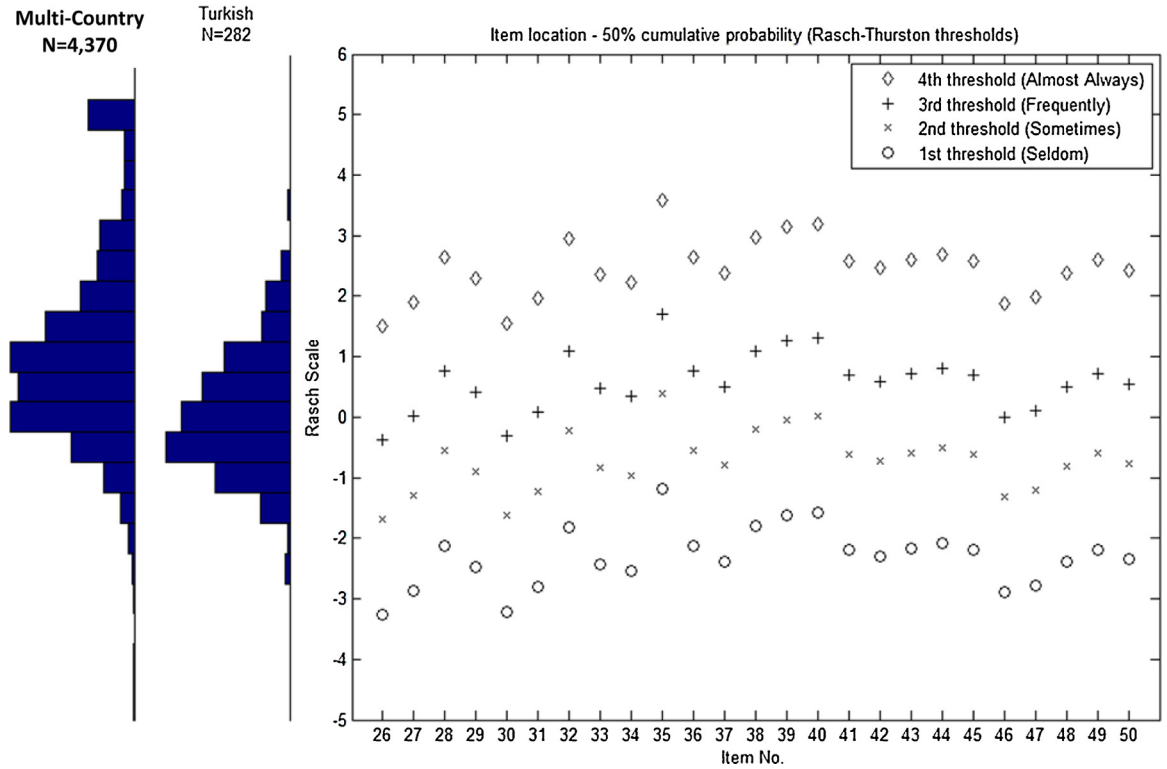

Fig. 10. Wright Map comparing Turkish and multi-country data for develops a positive school learning climate (PIMRS Teacher Form).

their subscales. We conclude that the PIMRS Turkish Form meets high standards of reliability and internal validity (American Educational Research Association (AERA), 1999).

The national profile suggest that Turkish principals engaged most actively in practices associated with the dimension, *Defines the School's Mission* dimension and least actively in practices related to the dimension *Manages the Instructional Program*. When we compared the instructional leadership profile of Turkish principals with a multi-national comparison group, the level of engagement among the Turkish principals in this role was lower than the six comparison societies across all construct measures. This suggests a need for further examination of the instructional leadership practices of Turkish principals. If additional research supports this finding, it could indicate a need for policy revision and capacity development aimed at strengthening the instructional leadership of primary school principals in Turkey.

5.2. Implications for research and practice

Educational administration is undergoing rapid change from a discipline dominated by scholarship generated largely in a few 'Anglo-centric' societies into a global discipline (Bush & Jackson, 2002; Hallinger & Leithwood, 1996; Leithwood & Day, 2007; Walker and Dimmock, 2002). This transformation of the field is challenging the capacity of scholars to engage in building and testing theories across a much more diverse set of national contexts than has been the case in the past (Belchetz & Leithwood, 2007; Hallinger & Leithwood, 1996; Lee & Hallinger, 2012; Leithwood & Day, 2007; Walker and Dimmock, 2002). Inherent in this challenge is the need for systematic development and validation of research tools that can be used across diverse societies (Hallinger and Bryant, 2013). The current study responded to the challenge of validating a widely used North American research instrument for use in Turkish policy research on educational leadership and management.

A practical result of this study lies in a validated Turkish Form of the PIMRS (Teacher Form). Although this study was limited to the validation of the PIMRS Teacher Form for use with Turkish primary school principals, we suggest that the research lays a strong foundation for broader use of the scale in Turkey. For example, future studies could easily employ the PIMRS Teacher Form (Turkish) as the basis for validating a PIMRS Principal Form (Turkish). Both instruments could be used in studies of secondary school principals as well as primary school principals. Although prior studies suggest that measurement results will be quite similar (Hallinger and Wang, 2015), scholars are advised to replicate the validation procedures employed in this study for confirmation.

We envisage Turkish scholars using the PIMRS-Turkish Form for national as well as cross-national research on principal instructional leadership. This would mirror, in many respects, the international program of research on leadership for learning (see Blasé, 1987; Hallinger & Heck, 1996; Heck & Hallinger, 2010, 2014; Leithwood & Day, 2007; Leithwood et al., 2008; Robinson, 2006). Relevant topics for study might include the investigation of principal leadership effects on teachers (e.g., teaching quality, job commitment, collective efficacy) as well as students (e.g., learning outcomes, engagement with the school).

The current study's initial attempt to build a 'national profile' of principal instructional leadership also has implications for policy research in Turkey. In Thailand this type of national profile was used to assess the impact of education reform on principal engagement in their instructional leadership role over time (see Hallinger & Lee, 2013, 2014). Policymakers in Turkey could use findings from the current study as a baseline measure of principal capacity in this domain and then evaluate the efficacy of future interventions related to principal preparation, selection and evaluation.

In terms of practice, increasing international interest in principal performance appraisal has raised the bar in terms of demonstrating the validity and reliability of research tools used in school systems (e.g., Catano & Stronge, 2006; Clifford & Ross, 2011; Fromm et al., In press; Goldring et al., 2009; Hallinger & Wang, 2015; Porter et al., 2010). The current report contributes to this developing literature by providing robust data concerning the reliability and validity of the Turkish form of the PIMRS. Thus, the PIMRS Turkish Form could be employed in principal needs assessment as well as performance appraisal when used in concert with other data sources (see Hallinger, 2012, 2014). Results of the current study do indicate a potential need for greater attention to developing the capacity of Turkish principals in managing the instructional program.

In conclusion, we join other scholars in acknowledging the importance of high quality research methods in the study of educational administration (Bridges, 1982; Hallinger, 2011; Hallinger and Heck, 1996). Indeed, we suggest that this task takes on even greater urgency in the context of the field's recent and rapid transformation into a global discipline. The past decade has witnessed a global expansion in the volume of research generated internationally. Unfortunately, the quality of research output undertaken outside of the traditional centers of scholarship in educational administration too often fails to meet 'international standards' (Hallinger and Bryant, 2013). This assertion has been affirmed in conversations with the editors of six international journals in educational administration.[3]

It remains an urgent priority to build a valid knowledge base in our discipline, one that reflects the diversity of school leadership practice across different national contexts. This depends on expanding the volume and geographical scope of high quality research. Achieving the dual targets of research quantity and quality continues to require thoughtful development and application of valid research tools. The authors hope that the current study offers a useful illustration of this principle in practice as well as possible directions for the application of research.

[3] The author has had occasion in Board meetings of these journals to discuss this issue in depth with editors as well as members of editorial boards of these journals.

Acknowledgement

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) for portions of this research.

References

AITSL (2014). *Australian professional standard for principals and the leadership profiles*. Victoria: Australia Institute for Teaching and School Leadership.
American Educational Research Association (AERA) (1999). *Standards for educational and psychological tests. American Psychological Association & National Council on Measurement in Education*. Washington, D.C: American Educational Research Association.
Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573.
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, Netherlands: Springer.
Blasé, J. (1987). Dimensions of effective school leadership: the teacher's perspective. *American Educational Research Journal*, *24*(4), 589–610.
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. (1982). Research on the school administrator: the state-of-the-art, 1967–1980. *Educational Administration Quarterly*, *18*(3), 12–33.
Brislin, R. W. (1986). Field methods in cross-cultural research. In W. Lonner, & J. W. Berry (Eds.), *Cross-cultural research and methodology series:* (Vol. 8. pp. 137–164).Sage: Thousand Oaks, CA.
Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.
Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: lessons from Chicago*. Chicago and London: The University of Chicago Press.
Bush, T., & Jackson, D. (2002). *A preparation for school leadership international perspectives*, *Vol. 30 (4*, Educational Management Administration & Leadership417–429.
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.
Clifford, M., & Ross, S. (2011). *Designing principal evaluation systems: Research to guide decision-making*. Washington, D.C: National Association for Elementary School Principals Accessed 14.01.15.
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: Wiley.
Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools*. Albany, NY: SUNY Press.
Day, C. (2009). Capacity building through layered leadership: Sustaining the turnaround. In A. Harris (Ed.), *Distributed leadership*Buckingham: Open University Press.
Duke, D. L., & Salmonowicz, M. (2010). Key decisions of a first-year 'turnaround' principal. *Educational Management Administration & Leadership*, *38 (1*, 33–58.
Erickson, D. A. (1979). Research on educational administration: the state-of-the-art. *Educational Researcher*, *8*(3), 9–14.
Firestone, W., & Herriot, R. (1982). Prescriptions for effective elementary schools don't fit secondary schools. *Educational Leadership*, *40*, 51–53.
Fromm, G., Volante, P., Hallinger, P., & Wang, W.C. (in press). A validation study and cross-cultural comparison of principal instructional leadership using the PIMRS Spanish language form. *Educational Management Administration & Leadership*.
Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., et al. (2009). The evaluation of principals: what and how do states and urban districts assess leadership? *The Elementary School Journal*, *110*(1), 19–39.
Hallinger, P. (2012). A data-driven approach to assess and develop instructional leadership with the PIMRS. In J. Shen (Ed.), *Tools for improving principals' work* (pp. 47–69).New York, NY: Peter Lang Publishing.
Hallinger, P. (2011). 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. (2014). *Measurement properties of the principal instructional management rating scale. Technical report (ver. 5.12)*. Bangkok, Thailand: Leading Development Associates.
Hallinger, P. (1982). *Principal instructional management rating scale*. Bangkok, Thailand: Leading Development Associates.
Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal*, *96*(5), 527–549.
Hallinger, P., & Bryant, D. A. (2013). Review of research publications on educational leadership and management in Asia: a comparative analysis of three regions. *Oxford Review of Education*, *39*(3), 307–328.
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. (2013). Exploring principal capacity to lead reform of teaching and learning quality in Thailand. *International Journal of Educational Development*, *33*, 305–315.
Hallinger, P., & Lee, M. S. (2014). Mapping instructional leadership in Thailand: has education reform impacted principal practice? *Educational Management, Administration and Leadership*, *42*(1), 6–29.
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 management behavior of principals. *Elementary School Journal*, *86*(2), 217–247.
Hallinger, P., Wang, W. C., & Chen, C. W. (2013). Assessing the measurement properties of the principal instructional management rating scale: a meta-analysis of reliability studies. *Educational Administration Quarterly*, *49*(2), 272–309.
Hallinger, P., & Wang, W. C. (2015). *Assessing leadership for learning with the principal instructional management rating scale*. Dordrecht, Netherlands: Springer.
Heck, R. H., & Hallinger, P. (2014). Modeling the effects of school leadership on teaching and learning over time. *Journal of Educational Administration*, *52*(5), 653–681.
Heck, R. H., & Hallinger, P. (2010). Testing a longitudinal model of distributed leadership effects on school improvement. *The Leadership Quarterly*, *21*, 867–885.
Heck, R. H., Larsen, T. J., & Marcoulides, G. A. (1990). Instructional leadership and school achievement: validation of a causal mode. *Educational Administration Quarterly*, *26*(2), 94–125.
Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
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.
Lang, G., & Heiss, G. D. (1998). *A Practical guide to research methods*. New York: 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*(4), 461–482.
Leithwood, K., & Day, C. (2007). *Successful school leadership in times of change*. Toronto, Ont: Springer.

Leithwood, K. A., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership and Management*, *28*(1), 27–42.
Leithwood, K. A., Harris, A., & Strauss, T. (2010). *Leading school turnaround: how successful leaders transform low-performing schools*. San Francisco, CA: Jossey-Bass.
Leithwood, K. A., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences learning. *Educational Administration Quarterly*, *46*(5), 671–706.
Leithwood, K. A., Louis, K., Anderson, S., & Wahlstrom, K. (2004). *Executive summary: how leadership influences student learning*. New York, NY: The Wallace Foundation.
Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*, 437–448.
Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: integrating transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370–397.
Murphy, J. F. (2007). Restructuring through learning-focused leadership. In H. J. Walberg (Ed.), *Handbook on restructuring and substantial school improvement* (pp. 63–75).Charlotte, NC: Information Age.
Murphy, J. F. (2008). *Turning around failing schools: leadership lessons from the organizational sciences*. Thousand Oaks, CA: Corwin Press.
Muthén, L. K., & Muthén, B. O. (2010). *Mplus 6*. Los Angeles, CA: Muthén and Muthén.
Neumerski, C. M. (2013). Rethinking instructional leadership, a review: what do we know about principal, teacher, and coach instructional leadership, and where should we go from here? *Educational Administration Quarterly*, *49*(2), 310–347.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, 3rd ed. New York: McGraw-Hill.
Porter, A. C., Polikoff, M. S., Goldring, E., Murphy, J., Elliott, S. N., et al. (2010). Developing a psychometrically sound assessment of school leadership: the VAL-ED as a case study. *Educational Administration Quarterly*, *46*(2), 135–173.
Rigby, J. G. (2014). Three logics of instructional leadership. *Educational Administration Quarterly* doi: 0013161X13509379.
Robinson, V. M. (2006). Putting education back into educational leadership. *Leading and Managing*, *12*(1), 62–75.
Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: a study of mediated pathways to learning. *Educational Administration Quarterly*, *48*(4), 626–663.
Spillane, J. P. (2005). Primary school leadership practice: how the subject matters. *School Leadership and Management*, *25*(4), 383–397.
Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*, 893–898.
Walker, A. D., & Dimmock, C. (2002). Moving school leadership beyond its narrow boundaries: developing a cross-cultural approach. In K. Leithwood, & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 67–204).Dordrecht, Netherlands: Kluwer.
Walker, A. D., & Ko, J. (2011). Principal leadership in an era of accountability: a perspective from the Hong Kong context. *School Leadership & Management*, *31*(4), 369–392.
Walker, A., & Kwan, P. (2012). Principal selection panels: strategies, preferences and perceptions. *Journal of Educational Administration*, *50*(2), 188–205.
Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement: the elusive search for an association. *Educational Administration Quarterly*, *34*(3), 398–425.
Wright, B., & Master, G. (1982). *Rating scale analysis*. Chicago, IL: MESA.
Wright, B. D., Linacre, J. M., Gustafen, J. E., & Martin-Lof, P. (1994). Reasonable meansquare fit values. *Rasch Measurement Transactions*, *8*, 370.