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

Assessing instructional leadership: a longitudinal study of new principals

Gavin T.L. Brown and Constance Chai
Faculty of Education, The University of Auckland, Auckland, New Zealand

Received 8 July 2011
Revised 17 November 2011
31 December 2011
8 February 2012
9 February 2012
Accepted 12 February 2012

Abstract

Purpose – The purpose of this paper is to evaluate the psychometric properties of the Self Assessment of Leadership of Teaching and Learning (SALTAL) inventory, in conditions of repeated administration.

Design/methodology/approach – In 2006 and 2007, nearly all of New Zealand's newly-appointed school principals participated in an 18 month induction program (First Time Principals). The SALTAL self-report was administered in three waves (i.e. before FTP, after two residential courses, and at the end of the FTP) to two cohorts. This voluntary survey was completed all three times by 55 per cent ($n = 86$) and 44 per cent ($n = 85$) of 2006 and 2007 participants respectively. Multi-group confirmatory factor analysis evaluated the stability of the SALTAL factor structure for each of the six administrations. Longitudinal curve modeling evaluated the linear effect of time on SALTAL responses.

Findings – Responses to SALTAL were found to be statistically equivalent across all six administrations. The longitudinal model was statistically invariant between cohorts. Initial scores were inversely correlated with changes over time. Increased time had a significant effect on SALTAL scores.

Originality/value – The paper shows that the SALTAL has demonstrable stability in eliciting response in repeated administration and is useful for studying the impact of leadership development programs.

Keywords New Zealand, Schools, Principals, Self-assessment, Instructional leadership, First-time school leaders, Self-reported capacity, Longitudinal curve modelling

Paper type Research paper

New research on how educational leaders make an impact on student outcomes provides increasingly specific guidance about the relative impacts of different types of leadership practice. The conclusions of several recent reviews of the evidence on the direct and indirect effects of leadership on student outcomes all point to the importance of instructional leadership (Blase and Blase, 2000; Goldring *et al.*, 2009; Hallinger, 2011b; Leithwood *et al.*, 2004; Quinn, 2002; Robinson *et al.*, 2008a). Robinson *et al.* (2008a, p. 2) concluded that "the more leaders focus their relationships, their work, and their learning on the core business of teaching and learning, the greater their influence on student outcomes."

The term instructional, or learning-focussed leadership, embraces a number of leadership practices, including setting and communicating clear instructional goals and expectations; strategic resourcing of priority goals; overseeing and evaluating teaching and teachers; promoting and participating in teacher learning and development and creating an orderly environment that is safe for and supportive of both staff and students (Alig-Mielcarek and Hoy, 2005; Robinson *et al.*, 2008a).

The authors are grateful to Professor Viviane Robinson, The University of Auckland, for providing access to the SALTAL data and for assistance with earlier drafts of the paper.

Journal of Educational Administration
Vol. 50 No. 6, 2012
pp. 753-772

0957-8234
DOI 10.1108/09578231211264676

The recent evidence about the impact of instructional leadership, combined with policy imperatives to have all students succeed on intellectually challenging curricula, have resulted in a new emphasis on building the instructional leadership capacity of both principals and of more widely distributed leadership teams (Elmore, 2004; Nelson and Sassi, 2005; Pont *et al.*, 2008; Spillane *et al.*, 2003). Efforts to develop individuals and teams are not sustainable, however, if the organizational and policy environments in which educators work are not strongly aligned to this goal. Recent analyses suggest that, in the USA at least, there is considerable misalignment. Adams and Copland (2007) analyzed the principal licensing standards of 50 states in the USA and found that, while a learning focus was included in about 34 states, it was emphasized in only six, with considerably more emphasis typically being given to general organizational skills and knowledge such as mentoring, communicating, and managing constituencies. They concluded that "few states have taken the important step of crafting licensing policies that reflect a coherent learning-focused school leadership agenda" (Adams and Copland, 2007, p. 181).

In a similar analysis, a group of researchers at Vanderbilt University examined the instructional leadership emphasis of 66 leadership assessment instruments used by some or all the school districts in 17 states in the USA. They found a greater emphasis on instructional leadership, with 52 percent of all items coded in this category, "as compared with management (15%), relations with the external environment (9%), and personal leadership (22%)" (Goldring *et al.*, 2009, p. 24). Despite this apparent focus on instructional leadership, the authors were critical of the superficial nature of many of the assessments, describing them as treating the content to be assessed as "a mile wide and an inch deep" (Goldring *et al.*, 2009, p. 25).

The assessment of instructional leadership

If we are to monitor and evaluate the consequences of investment in the development of instructional leadership, we need assessment tools that are technically sound and strongly focussed on this type of leadership. Of the 66 leadership assessment tools analyzed by the Vanderbilt team, the vast majority "have limited or no published information concerning their reliability or validity" (Goldring *et al.*, 2009, p. 26). They concluded that their use for moderate to high stakes assessment decisions was in violation of professional testing standards (American Educational Research Association (AERA), American Psychological Association (APA), and National Council for Measurement in Education (NCME), 1999).

While the tools used by states and school districts to assess school leaders fall short on the criteria of strongly focussed on instructional leadership and technically sound, some more promising assessment tools can be found in the research literature on instructional leadership. The most well known is the Principal Instructional Management Rating Scale (PIMRS) which was developed by Hallinger in the early 1980s and has since been used in over 130 doctoral studies of instructional leadership (Hallinger, 2011a). The scale comprises 71 items organized into 11 subscales (i.e. frame school goals, communicate school goals, coordinate curriculum, supervise, and evaluate instruction, monitor student progress, protect instructional time, provide incentives for teachers, provide incentives for learning, promote professional development, and maintain high visibility). Each item describes a critical job-related behavior. Raters of the principal's behavior, who may be teachers, district superintendents, or the principals themselves, are asked to indicate the frequency with which the principal has demonstrated the specified behavior in the past year.

A second assessment tool has been developed by the US-based Educational Testing Service (ETS) that has a strong focus on the leadership of learning. The tool, known as the School Leaders Licensure Assessment (SLLA), is anchored in six standards (i.e. Standard 1 facilitates the development, articulation, implementation, and stewardship of a vision of learning shared and supported by the school community; Standard 2 advocates, nurtures, and sustains a school culture and instructional program conducive to student learning and staff professional growth; Standard 3 manages the organization, operations, and resources for a safe, efficient, and effective learning environment; Standard 4 collaborates with families and community members, responding to diverse community interests and needs, and mobilizing community resources; Standard 5 acts with integrity, fairness, and in an ethical manner; and Standard 6 understands, responds to, and influences the larger political, social, economic, legal, and cultural context) which express the broad strategies through which school leaders promote the educational success of their students (ETS, 2009). The standards were developed in the USA by the Interstate School Leaders Licensure Consortium (ISLLC) (1996), a coalition of state officials, professional associations, and academic groups (Council of Chief State School Officers (CCSSO), 2008; Murphy, 2003, 2005). The consortium hoped the standards would shift the field of educational administration from its preoccupation with generic business management and social science disciplines, to a much stronger focus on the knowledge, skills and dispositions that research evidence and professional expertise suggested were required to lead schools that delivered high quality and equitable outcomes for students.

Given the high stakes nature of much leadership assessment, it is important that assessment tools meet established psychometric standards. The PIMRS, for example, has been validated in multiple independent studies in which exploratory factor analysis and discriminant analysis were used to establish the reliability and validity of scores (Hallinger, 2011a). Scale reliabilities (i.e. Cronbach's α) for the multiple factors (determined with EFA) were always >0.75. The SLLA, based on data from $>20{,}000$ tests, reports only a total, aggregated transformed score (i.e. range 100-200) obtained from four sections of the test, with a standard error of measurement estimated to be 4.40 (ETS, 2009). This means the reported score is a reliable overall estimate of performance, but lacks precision as to dimensions of instructional leadership.

While the above tools have undergone considerable psychometric development, two threats to their validity remain. First, it may be that the conceptual model that is confirmed by the results of a particular validation may be particular to the sample on which the assessment tool was developed. While large samples provide resistance to chance artifacts, confirmation of a model with the same sample with which the model was first developed cannot successfully falsify the model. Cross-validation addresses this threat by administering the instrument to a new cohort and testing whether the model is confirmed. Such evidence would be found by testing if the preferred model varied by no more than chance for different groups (i.e. multiple-group confirmatory factor analysis invariance) (Byrne *et al.*, 1989; Marsh and Hocevar, 1985; Vandenberg and Lance, 2000).

A second threat arises from the fact that validation studies that employ a single administration of an instrument do not establish a strong warrant for using it in longitudinal or repeated measures studies. It may be that data from a single cohort administration will not generalize to multiple administrations (i.e. there may be a practice or fatigue effect). Hence, the stability of the factor structure should also be examined under a repeated measures condition, especially if it is intended that the

instrument be used to monitor changes in the capability of respondents. If the original measurement model is replicated at each time point, then evidence is provided for the consistency and stability of the tool across repeated measurements. Thus, ideally the development of leadership inventories would combine repeated measures and independent samples. This paper reports a study in which a self-assessment of instructional leadership was validated in this way.

Building instructional leadership capacity in New Zealand

The drive to develop instructional leadership capacity and ways to measure it has not been confined to the USA. Parallel developments are occurring throughout the OECD (Pont *et al.*, 2008). In New Zealand, the research confirming the impact of instructional leadership on student outcomes has been widely disseminated among educational leaders and policy makers and shaped the policy framework which guides all leadership development initiatives in New Zealand (Ministry of Education, 2008).

In New Zealand, responsibility for induction of school principals as leaders of learning and teaching lies mostly with the providers of the national induction program for newly appointed school principals. Known as the "first-time principals" (FTP) program, it consisted, at the time of this study, of nine days of residential training in three blocks of three days, mentoring, on-line learning, and research and evaluation. A major focus of the research strand has been the development and validation of a tool to assess the capability of the new appointee in the leadership of teaching and learning.

The tool, known as the Self-Assessment of the Leadership of Teaching and Learning (SALTAL), comprises 24 items organized into four dimensions (items are provided in Appendix). A review of the research literature on leadership effects (Robinson, 2007) partially informed the development of the SALTAL. The identified most powerful effects on student learning revolved around leaders who promoted and participated in teacher learning (SALTAL Dimensions 1 and 4), planning, coordinating, and evaluating teaching (SALTAL Dimension 3), establishing goals and expectations (SALTAL Dimension 2), strategic resourcing, and ensuring an orderly and supportive environment. Dimension 1, "knowledge and skills for leading teaching and learning," assesses the extent to which the principal has a deep understanding of the relationship between teaching, learning, and leadership. Dimension 2, "commitment to ensuring positive learning outcomes for all students," assesses the extent to which the principal focusses on and prioritizes the core business of student learning. Dimension 3, "collaborative leadership," assesses the extent to which the principal works through and with others to effectively lead and manage the operations of the school, mindful of the impact of decisions on students. Dimension 4, "ethical leadership," is defined as the extent to which the principal models personal and professional integrity. It includes aspects of the management of self and others, such as the management of workload and the ability to deal fairly and effectively with tough staff issues. Principals rate on a five-point scale the extent to which they believe they currently demonstrate the aspect of leadership described by each of the 24 items. Each item is accompanied by three or four indicators which provide concrete illustrations of exactly what is meant by each item.

The need to develop both items and indicators that had high face validity for instructional leadership activities and which were situated in the New Zealand educational administration context was the major reason for developing a context-specific measure of instructional leadership, rather than adopting previously published assessment tools. Nonetheless, there is considerable overlap in conceptual

framing of the nature of instructional leadership among the three assessments (Table I). The inventories share a common understanding that leadership has to concern itself with instructional programming, school climate, and collaboration with staff and community. There is considerably more detail in both the PIMRS and SLLA approaches, relative to SALTAL. While it is appropriate for leaders to be concerned for the many faceted aspects of school administration and management identified in PIMRS and SLLA, the focus of SALTAL is on activities directly related to improving student learning outcomes. Hence, the use of SALTAL is highly useful when data collection is motivated by an interest in learning-related activities, as opposed to general school administration.

Furthermore, the SALTAL, unlike PIMRS and SLLA, is a self-evaluation tool. Self-reporting is considerably less powerful than external observations of behaviors in determining whether instructional leadership practices are being implemented since the technique is subject to biases (Donaldson and Grant-Vallone, 2002) such as self-enhancing tendencies (e.g. not reporting aspects which make the individual feel bad about themselves) and social desirability tendencies (e.g. reporting gains in order to please funders or professional development providers). It has been found that school principal self-ratings were optimistic relative to the ratings provided by their teachers, which were closer to independent sources about the principal performance (Hallinger, 2011a). Nonetheless, self-other (e.g. principal-teacher) correlations as to characteristics of the self have been found to be only moderate (mean $r = 0.395$) (Kenny and West, 2010) suggesting there might be no inherent superiority in either the self or the other in

SALTAL (Robinson *et al.*, 2006, 2008a, b)	PIMRS (Hallinger, 2011a)	SLLA (CCSSO, 2008)
D1: knowledge and skills for leading teaching and learning	*Manage instructional program*: coordinate curriculum Supervise and evaluate instruction Monitor Student Progress	*Standard 3*: manages the organization, operations, and resources for a safe, efficient, and effective learning environment *Standard 6*: understands, responds to, and influences the larger political, social, economic, legal, and cultural context
D2: commitment to ensuring positive learning outcomes for all students	*Defining school mission*: frame school goals Communicate school goals	*Standard 1*: facilitates the development, articulation, implementation, and stewardship of a vision of learning shared and supported by the school community *Standard 4*: collaborates with families and community members, responding to diverse community interests and needs, and mobilizing community resources
D3: collaborative leadership *D4*: ethical leadership	*Develop school learning climate program* Protect instructional time Provide incentives for teachers Provide incentives for learning Promote professional development Maintain high visibility	*Standard 2*: advocates, nurtures, and sustains a school culture and instructional program conducive to student learning and staff professional growth *Standard 5*: acts with integrity, fairness, and in an ethical manner

Table I.
Content mapping of three school leadership inventories

evaluating a leader's performance. Factors that impact possible agreement between self- and other-ratings influence both parties in the equation and include sex, age, race, education, personality, metacognitive ability, and cultural context (Fleenor *et al.*, 2010). Hence, while triangulation between the principal and relevant others in the school is advised, it is still potentially useful to examine changes in the principal's self-rating, especially in the context of a professional development program.

Nonetheless, structured self-reporting mechanisms provide access to the internal opinions, attitudes, and beliefs of the individual. This information provides useful data for understanding participant behavior since attitudes, beliefs, and opinions are known predictors of intentions and actions (Ajzen, 2005). Furthermore, the act of reporting one's own opinions contributes to the individual's own personal reflection, which may help with further development of the individual school leader. It should also be noted that in evaluating any phenomenon there are strong methodological effects; in other words, how data are gathered shapes the type of results discovered. For example, a review of interview and questionnaire data collection around common constructs (Harris and Brown, 2010) found that, at best, the level of agreement between methods meant that results would be complementary rather than confirmatory. Consequently, the use of a self-report inventory is a useful adjunct in a multi-method approach to determining what changes, if any, are occurring in school leaders.

The approach taken in the SALTAL is an analytic one in which participants judged themselves item by item rather than make a holistic self-evaluation. A recognized virtue of this approach is that it generates a profile analysis that means strengths and weaknesses of an individual can be identified; whereas, a holistic approach may obscure problems or success within an overall rating. A further strength to the analytic approach, with its randomized presentation of items, is that the tendency to rate oneself in one way for all related items is weakened since they are not presented together. Yet another virtue to the analytic approach, common to all psychometric tests and inventories, is that it reduces the error inherent in any single rating. By having participants complete multiple items around a common construct, the true value for that construct is more likely to be found and is less prone to error inherent in a single measure of the construct.

Thus, notwithstanding issues to do with self-report, evidence of behavioral validity has been found. The theoretical and empirical origins of SALTAL, along with its psychometric properties, have been described in two earlier publications (Robinson *et al.*, 2006, 2008b). Our purpose in this paper is to investigate the psychometric properties of the tool as a repeated measure. The two validity studies already published about SALTAL involved assessment of a cohort of principals at one point in time. In a study of 261 FTPs from the 2004 and 2005 cohorts, the four-factor structure of the SALTAL was established through exploratory factor analysis (Robinson *et al.*, 2006). Estimates of reliability for each factor were acceptable to strong (i.e. Cronbach's $\alpha = 0.69$, 0.81, 0.85, and 0.89). Further validation was done by determining the extent of the relationship between principals' self-assessments and those of their experienced principal mentors. Discriminant analysis of the SALTAL self-assessment scores assigned between 47 and 70 percent of FTPs to the same category of development need (high, medium, or low) as their mentors, thus providing validity evidence that SALTAL self-ratings were moderately to strongly aligned with each other. SALTAL responses from 121 FTP participants in the 2006 cohort replicated the earlier scale reliabilities (Robinson *et al.*, 2008b). Discriminant analysis of the SALTAL scores accurately assigned 70 percent of principals to the same leadership need classification as expert

mentors. In combination, this evidence shows that SALTAL is a valid and robust self-rating of instructional leadership for use with new principals.

Research questions and hypotheses

The purpose of this paper was to ascertain the psychometric properties of SALTAL when used as a repeated measure – administered three times – over the 18-month period of the FTP program. One step in validating a measurement tool is to demonstrate that it elicits consistent behavior no matter how frequently it is used. In the case of a self-report tool, it is important that each item belongs to the same factor regardless of frequency or timing of administration. This does not mean that people give the same score value at each instance of administration or that the correlation between factors remains constant. This requirement is based on the logical validity consideration that the items are a function of a distinct and consistent latent trait, not characteristics of the context in which the data are collected. Thus, *H1* expects that responses to the SALTAL will consistently aggregate into the four separate dimensions previously identified rather than any other structure (e.g. one generic leadership construct). Evidence for this hypothesis would be obtained if the structure of the SALTAL as four identifiable dimensions would be statistically invariant in repeated administrations within a cohort of FTP school leaders and invariant in a second independent cohort of first-time school leaders.

When data are collected repeatedly it is possible that responses at subsequent times are solely dependent on starting values or are solely a reflection of circumstances present at the time of data collection. Validation evidence for a repeated measure, derived from the logic of longitudinal curve modeling, comes about when it is clear that there is an effect on responding due to initial values and time-related changes in circumstances. Thus, *H2* expects that participant responses to the SALTAL over the three waves of administration would be a function of both initial starting points and changes related to time. The hypothesis is strengthened if the same pattern is replicated in an independent cohort of participants.

Method

These hypotheses were tested with a sample drawn from the 2006 cohort of the First-Time Principals Project conducted in New Zealand. A second sample from the (2007) cohort was used to cross-validate the model.

Participants

There were 155 first time principals (92 female and 63 male) in the 2006 cohort used to develop the model of instructional leadership that became the SALTAL instrument. Of these 155 principals, 86 (58 female and 28 male) or 55 percent completed all three administrations of SALTAL. Similarly, there were 197 principals in the 2007 cohort, of whom 86 (56 female and 30 male) or 44 percent completed SALTAL at all three times. Table II provides a description of some personal and school characteristics of the two samples. There were no significant differences between the full cohort and the samples on any of the variables.

Participants were predominantly New Zealand European, aged over 40, with more than ten years experience, and held a bachelors degree or higher. The majority of FTPs in both samples led elementary schools and there was a reasonably equal representation of schools from low, middle, and high socio-economic communities.

Personal and school characteristics	2006 ($n = 86$)		2007	
	n	%	*n*	%
Personal characteristics				
Ethnicity				
New Zealand European	68	79	69	80
Māori	12	14	10	12
Other	6	6	7	8
Age (years)				
21-30	7	8	4	5
31-40	19	22	32	37
41-50	40	47	35	41
51 +	20	23	15	17
Years of teaching experience (years)				
1-5	7	8	9	11
6-10	9	10	20	23
11-15	23	27	17	20
16-20	17	20	13	15
21 +	30	35	27	31
Highest qualification				
Below degree level	30	35	23	43
Bachelor's degree	38	44	43	50
Graduate/post graduate diploma	9	10	10	12
Master's degree or higher	9	10	10	12
School characteristics				
School type				
Elementary	64	74	73	85
Middle	5	6	3	3
Secondary	13	15	8	9
Māori medium	3	3	1	1
Special	1	1	1	1
Socioeconomic status				
Low	28	33	29	34
Middle	40	47	34	40
High	18	21	23	26

Table II.
Personal and school characteristics for two samples of first-time principals

Instrument: the self-assessment of leadership of teaching and learning

The SALTAL comprises 24 items, grouped according to four key leadership dimensions. These dimensions, along with the psychometric properties of the instrument, have already been described. Each dimension is assessed with between four and nine items. Participants respond using a five-point rating scale to indicate the extent to which they demonstrate the indicated aspect of leadership (i.e. to what degree do you …), a 1 indicating that the aspect is present "hardly at all" and a 5 that it is present to a "high level." As soon as participants enrolled in the induction program they were asked to complete their first SALTAL assessment. This initial baseline assessment was followed by a second administration that took place at the end of the second residential course. The final SALTAL data were collected after the third and final residential course.

Analysis

H1 was tested using nested, multi-group confirmatory factor analysis and *H2* was evaluated with longitudinal curve modeling of the SALTAL scale scores. Two cohorts

(i.e. 2006 and 2007) provided three waves or panels of repeated measures collected at occasions fixed relative to the residential courses over an 18-month period. Since the same instrument was used at each time the data are on comparable scales. Missing data points (i.e. only 3.2 percent of all data points in 2006; 0.4 percent in 2007) were imputed with the expectation maximization procedure (Dempster *et al.*, 1977):

H1. Invariance of SALTAL across repeated administrations.

The psychometric properties of the SALTAL at each time of administration for both cohorts were tested with nested, multi-group confirmatory factor analysis. This meant that one analysis of the model was run using six groups (i.e. Time 1, Time 2, and Time 3 for each cohort). The invariance of a model across the separate times is evaluated by establishing whether the differences in model parameters (e.g. factor regression weights, covariances, factor intercepts, or residuals) are statistically significant (Cheung and Rensvold, 2002; Vandenberg and Lance, 2000), strict invariance (i.e. equivalent residuals) is not necessary; whereas configural and metric invariance (i.e. all zero parameters and all factor regression weights) must be present in order to use factor scores in longitudinal analysis (McArdle, 2007), The difference in χ^2 taking into account the difference in df should not be statistically significant (i.e. $p \geqslant 0.05$) and/or the difference in CFI should be $\leqslant 0.01$ for invariance to be demonstrated.

A complication arising from the small sample size when testing the stability of the full 24-item, four-factor SALTAL model in this study was the low ratio of cases to variables (i.e. 86 participants to 24 items = 3.6:1). In these conditions, it is possible that mathematically >100 percent of the variance in a factor could be explained by the model (Chen *et al.*, 2001). This negative error variance can be set to a small positive value (e.g. 0.005) if there is evidence that when the ratio of cases to variables is more robust (e.g. 10:1) the observed value of the variance is >0.00. Additionally, if twice the standard error of the observed negative error is larger than the error value then it is highly probable that the real variance value is >0.00. Negative error variances meeting these two conditions were fixed to a small positive value in this study:

H2. Change in SALTAL responses over time.

Examining how participants respond to an inventory across repeated administrations can establish both the validity and reliability of an inventory. It is now conventional to analyze repeated measures using a latent curve model (LCM) approach that estimates the time-varying effect of an initial intercept (i.e. starting score) parameter and a slope (i.e. rate of change) parameter on the repeated scores (Bollen and Curran, 2006). However, this approach may not be estimable with sample sizes below 100, which applies in this study. A solution to small sample size is to parcel manifest variables contributing to latent factors into single scores (Little *et al.*, 2002). Aggregation into scales is warranted if there is robust evidence that the items pool into the hypothesized factors at each time of administration. In this study, the longitudinal curve model treated the SALTAL as a single factor consisting of four parceled manifest variables, repeated three times with equal intervals of about nine months between data points.

Results

Table III provides the SALTAL dimension inter-correlation matrix for all cases aggregated across 2006 and 2007. The scale estimates of internal reliability are robust

	Time 1				Time 2				Time 3			
	D1	D2	D3	D4	D1	D2	D3	D4	D1	D2	D3	D4
Time 1												
D1: knowledge and skill for leading teaching and learning	0.70											
D2: commitment to ensuring positive learning outcomes for all students	**0.65**	0.84										
D3: collaborative leadership	**0.62**	**0.75**	0.87									
D4: ethical leadership	**0.52**	**0.67**	**0.72**	0.79								
Time 2												
D1: knowledge and skill for leading teaching and learning	*0.55*	0.38	0.30	0.30	0.73							
D2: commitment to ensuring positive learning outcomes for all students	0.35	*0.43*	0.29	0.30	**0.62**	0.85						
D3: collaborative leadership	0.40	0.43	*0.48*	0.41	**0.67**	**0.74**	0.87					
D4: ethical leadership	0.26	0.33	0.33	*0.47*	**0.52**	**0.64**	**0.72**	0.74				
Time 3												
D1: knowledge and skill for leading teaching and learning	*0.36*	0.24	0.20	0.17	*0.52*	0.34	0.38	0.32	0.70			
D2: commitment to ensuring positive learning outcomes for all students	0.23	*0.26*	0.22	0.15	0.31	*0.46*	0.38	0.32	**0.58**	0.80		
D3: collaborative leadership	0.17	0.17	*0.18*	0.20	0.33	0.37	*0.51*	0.42	**0.61**	**0.63**	0.86	
D4: ethical leadership	0.16	0.21	0.19	*0.24*	0.29	0.31	0.38	*0.48*	**0.53**	**0.59**	**0.70**	0.75

Table III.
SALTAL dimensions inter-correlations across three administrations (aggregated data 2006 and 2007)

Notes: $n = 171$; values on diagonal are Cronbach α scale estimates of reliability from 2006 data only; values in bold are within time, across dimensions inter-correlations; values in italics are within-trait, across time correlations; all reported values are statistically significant

with a mean of $\alpha = 0.79$ indicating strong inter-correlation of items within each scale. The mean inter-correlation at each time across the dimensions was $r = 0.64$, indicating moderate within-time consistency among the dimensions. The mean inter-correlation of the same dimension across time was $r = 0.41$, suggesting a weak tendency for scores in each dimension to be consistent across times of administration. Nonetheless, the across-time within-dimension correlations were considerably weaker between Times 1 and 3 (mean $r = 0.26$) than between the contiguous times, suggesting an autoregressive characteristic (i.e. meaning that the past predicts the present, but the influence decreases over time) in dimension traits (*H1*).

The 24-item, four-factor hierarchical model identified in previous studies was tested six times (i.e. three times per cohort) for fit using confirmatory factor analysis. Multi-group invariance testing examined whether the parameters were identical in each of these six models. Negative error variance was found once in only one dimension for only one of the six times (i.e. 2007 cohort, Time 1, Dimension 2). Because the standard error (0.012) was greater than the negative error value (−0.011) and because all other models had shown positive variance, the offending value was fixed to 0.005. Consequently, the unconstrained model for the six group invariance test (Figure 1) had good fit to the data ($\chi^2 = 2266.91$; df = 1494; χ^2/df = 1.52, $p = 0.22$; CFI = 0.84; gamma hat = 0.98; SRMR = 0.075; RMSEA = 0.032).

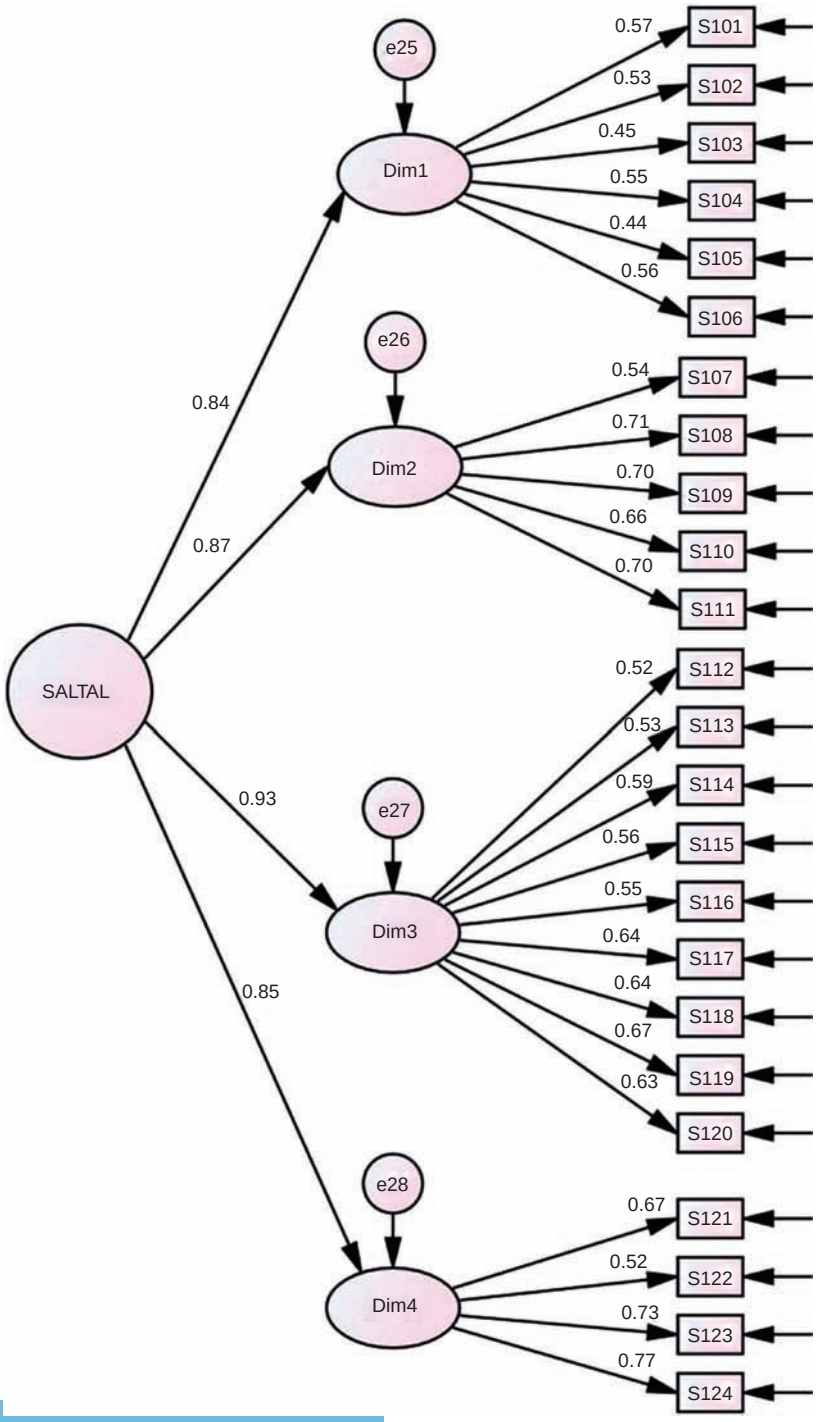

Figure 1.
Full factor measurement model of SALTAL inventory aggregated for six administrations

Sequential testing of the progressively constrained model found that, except for residual values, the change in CFI was <0.01 and the change in the χ^2 had a statistically non-significant p value. This indicated that all paths were identical (i.e. configural invariance) and that all regression weights (i.e. metric invariance), and all covariances between factors were equivalent (i.e. scalar invariance) (Table IV). In other words, any differences in the SALTAL parameters over time represent differences due to chance factors rather than differences in how the participants responded to the inventory items. This finding supports the conclusion that the inventory functions in a stable fashion across time and groups of participants and that the participants are all members of the one population of first-time school leaders (Wu *et al.*, 2007) (*H2*).

Since *H1* analysis showed that FTP responses to the SALTAL were statistically equivalent across all six times of administration, there was a strong warrant to consider that the two cohorts could be treated as one group and that the average of items for each SALTAL dimension would be a good summary of responses. Hence, to examine longitudinal effects, an averaged, parcel variable for each factor was created for each time of administration. This resulted in 12 parceled manifest variables for analysis in a longitudinal curve model. Parceling gave a sample to variable ratio of 171:12 (i.e. 14.25:1), well exceeding minimum recommended ratios for structural equation models (Costello and Osborne, 2005). To take advantage of the larger sample size, a longitudinal model was created based on all 171 participants as if they had provided data simultaneously. The validity of the model was further tested by examining its equivalence for the two cohorts separately. Statistical equivalence of the longitudinal model across the two cohorts would mean that a single model was sufficient for both groups.

The LCM model (Figure 2) consisted of a single factor (SALTAL) predicting scores on the four averaged dimension scores repeated three times. To account for time, two correlated latent traits (i.e. mean and slope) were introduced to account for initial starting values and change in responses over time. To further account for the inter-correlation of each dimension with itself over time (Fitzmaurice *et al.*, 2004; Singer and Willett, 2003), the residuals of each dimension variable were correlated to itself only across the nearest time of administration (i.e. from Time 1 to Time 2 and from Time 2

Model	χ^2	$\Delta\chi^2$	df	Δdf	p	CFI	ΔCFI
Figure 1: SALTAL measurement model for six times of administration across 2006 and 2007							
Unconstrained baseline	2,266.91		1,494			0.839	
Equivalent measurement regressions	2,365.37	98.46	1,594	100	**0.53**	0.840	**0.001**
Equivalent structural regressions	2,382.41	17.04	1,609	15	**0.32**	0.839	**0.001**
Equivalent structural covariances	2,398.66	16.26	1,614	5	0.01	0.837	**0.002**
Equivalent structural residuals	2,520.57	121.91	1,630	16	<0.01	0.812	0.025
Equivalent measurement residuals	2,832.22	311.65	1,750	120	<0.01	0.787	0.025
Figure 2: LCM model of parceled SALTAL dimensions between 2006 and 2007 cohorts							
Unconstrained baseline	105.35		86			0.985	
Equivalent measurement regressions	114.90	9.55	95	9	**0.39**	0.984	**0.001**
Equivalent structural covariances	116.26	1.36	98	3	**0.72**	0.986	**0.002**
Equivalent structural residuals	120.17	3.91	121	3	**0.27**	0.985	**0.001**
Equivalent measurement residuals	175.28	55.11	121	20	<0.01	0.957	0.028

Table IV.
Nested invariance tests

Notes: CFI, comparative fit index; values marked in bold indicate statistical equivalence assumption holds; 2006 cohort $n = 86$; 2007 cohort $n = 86$

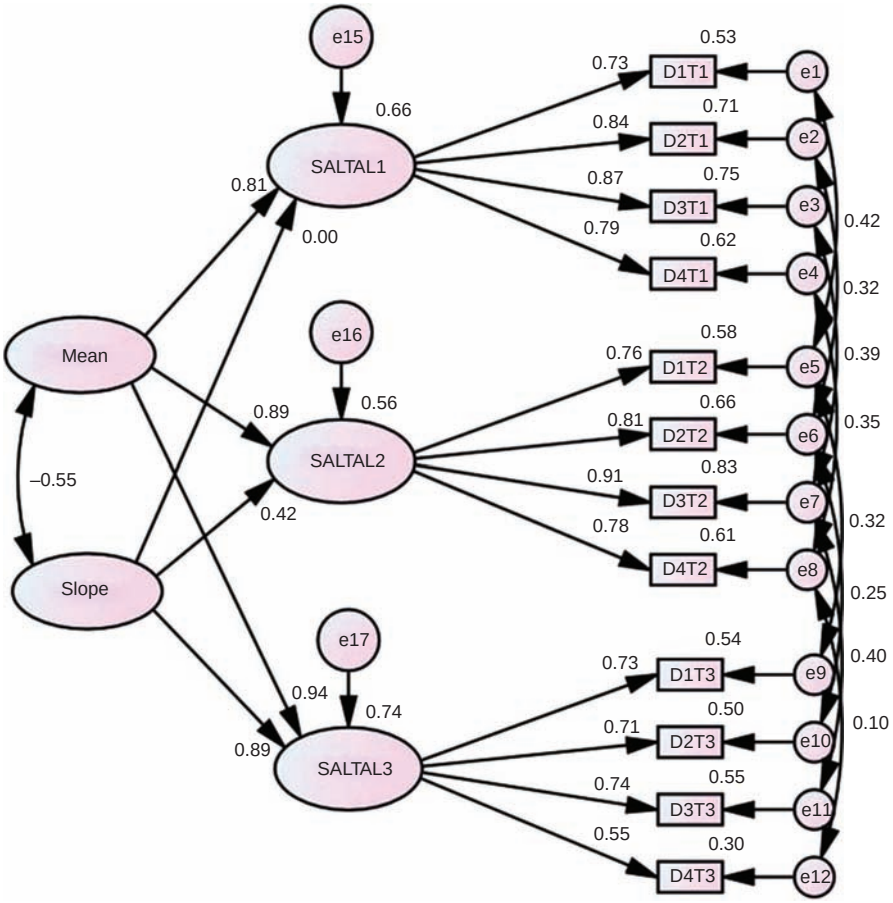

Figure 2.
LCM structural model for repeated administrations of the parcelled SALTAL inventory

to Time 3). Note that the correlation from Time 1 to Time 3 was not entered into the model, in accordance with the weak inter-correlations over the 18-month period shown in Table III. This model (Figure 2) had excellent fit to the data ($\chi^2 = 52.44$; df = 43; $p = 0.15$; CFI = 0.99; gamma hat = 0.99; SRMR = 0.038; RMSEA = 0.036).

To test the equivalence of the aggregated longitudinal curve model for the two cohorts (i.e. 2006 and 2007) a multi-group nested invariance test of Figure 2 was conducted. Table IV shows that the change in CFI was < 0.01 and the change in χ^2 had a statistically non-significant p value except when the measurement errors were constrained. This indicated that the two cohorts had identical path models (i.e. configural invariance) and statistically equivalent regression weights (i.e. metric invariance), and equivalent covariances among factors (i.e. scalar invariance). These equivalences indicated that observed differences between cohorts in responding to the SALTAL over time were attributable to chance and that the two samples were drawn from a single population of first-time principals. It also means that the single model reported in Figure 2 accounts for both groups satisfactorily.

The mean starting point and change slope were inversely correlated, indicating that increased scores in SALTAL across time tended to be associated with participants who

had lower starting scores, while those with higher initial scores had decreases. The starting mean had a relatively constant explanatory effect on the SALTAL responses ($\beta = 0.81$, 0.89 and 0.94, respectively). The change rate by the third time of administration had a nearly equal effect to the starting mean ($\beta = 0.89$), indicating that increased scores on the SALTAL inventory were just as much influenced by effects associated with the passage of time as by starting scores. Furthermore, the proportion of variance explained in the SALTAL scores at the third time of administration (SMC = 0.74) was more than the amount of variance explained at the start of the program (SMC = 0.66), suggesting time-based changes contributed to explaining participant responses to the SALTAL.

The residual inter-correlations became weaker in accordance with the autoregressive nature of the repeated measures (Time 1 to Time 2 mean $r = 0.37$; Time 2 to Time 3 mean $r = 0.27$). This suggests that the tendency to give the same response had a relatively weak contribution to changes in participant responses over time. The relatively smaller values of the residual inter-correlations also indicated that initial starting values and time-related changes were much more significant contributors to participant responses.

Discussion

This study has shown that the SALTAL factor structure is replicated across three administrations over an 18-month period in two cohorts of new school principals (*H1* accepted). The inventory is robust in how the items relate to the factors and in how the factors relate to each other. This makes the SALTAL a useful tool in identifying changes in self-perceptions of school principals concerning their instructional leadership capabilities. The stability of the SALTAL across six administrations in two groups indicates that any changes observed in the respondents are not attributable to deficiencies in the SALTAL inventory. The model underlying repeated administration of SALTAL is able to detect the extent to which participant responses are less than genuine. Specifically, the model can identify the degree to which participants simply repeat the responses they started with. This suggests that observed changes in responses to the SALTAL are likely to express some real world mechanism, rather than an aberration in the inventory.

The model has also shown that, as expected, there was a change in SALTAL scores as a function of time (*H2*). The LCM result supports the conclusion that SALTAL self-ratings at the end of the program are not purely a consequence of how principals rated themselves at the start of the program. As first-time principals gained more experience and received more feedback and guidance, their self-evaluations changed, potentially becoming either more optimistic or realistic depending on whether they were low or high to start with. There is evidence in the self-evaluation literature that with greater proficiency (Barnett and Hixon, 1997; Mitman and Lash, 1988; Sung *et al.*, 2010) and experience (Alsaker, 1989; Butler, 1990; Kasanen *et al.*, 2009; Wilson and Wright, 1993), self-ratings of ability tend to decrease, yet become more consistent with other measures of performance. Hence, it may be that these results indicate that less-confident principals gained in ability, while more-confident principals gained a more realistic self-perception as a consequence of their time as a first-time principal. However, it may be useful to consider, in future SALTAL studies, making the response scale longer than the current five points. If new hires tend to be unrealistically enthusiastic and optimistic about their new role, changes in their initial self-ratings may be obscured by the relatively low ceiling created by the scale. Rating scales that

are known to be more resistant to self-enhancement responding, such as positively packed rating scales (Klockars and Yamagishi, 1988; Lam and Klockars, 1982), could be considered.

While this study does not identify what observable differences there were in each principal's actual leadership practices, it does suggest that there were real changes in self-perceptions. Changes in self-evaluation are likely, in the context of professional development, to suggest more sophisticated understandings of the principal's own capabilities. One advantage of a self-report tool, notwithstanding acknowledged limitations of the methodology, is that it does permit reflection which can become the basis of conversations about needs, strengths, and opportunities between principals and developers. Of course, other validation techniques are required to eliminate the possibility that participants were only claiming improvements in accordance with the professional development program (a form of "teacher pleasing") or in accordance with their own investment in participating in the program (a form of "halo" effect).

As argued previously (Robinson *et al.*, 2008a, b) each of the four dimensions of the SALTAL is needed in eliciting self-ratings about leadership practices. The simple aggregation of items into a total scale would reduce the amount of information available to first-time principals themselves, their mentors, and/or professional development providers. Further, the presence of four dimensions of leadership practices persistently over time permits identification of where self-evaluated changes could be taking place – greater sensitivity to detect change in four different dimensions of instructional leadership increases the effectiveness of SALTAL.

This study is not able to establish the exact time or cause of the changes detected by the SALTAL. Exactly how much time is required to see change is still an open question; but it certainly seems to be longer than six months. A further confound in this study is that the unique effect of professional development components (e.g. mentoring, small groups, residential courses, lectures, etc.) and on-the-job experience cannot be determined. Further research is also required to identify whether the changes in SALTAL ratings result in better leadership. It is possible that principals whose SALTAL ratings changed (whether higher or lower) by a significant margin were better practitioners than those whose scores remained the same. Furthermore, studies which require participants to provide multiple instances of self-report data can be affected by unexplained withdrawal of participation; it is possible those who did not participate had quite different self-reports. Hence, future studies should seek to ensure 100 percent data completion by all participants.

Nonetheless, this study has provided a robust basis for accepting the validity of the SALTAL dimensions and for use of the assessment tool even when repeated administrations are required. This means the inventory can be useful frequently as a means of monitoring planned changes in principal practices; understanding educator belief systems is important since they play an important part in mediating how educational reforms are implemented in schools and classrooms (Richardson and Placier, 2001). Hence, giving participants, as well as developers, a way to consider their own beliefs, opinions, or attitudes about instructional practices is likely to be an effective adjunct to any deliberate attempt to improve the quality of instructional leadership.

Several further points about the SALTAL are warranted. The SALTAL is a useful addition to the relatively small set of available tools for assessing the role of the school principal as an instructional or learning leader, a matter of increasing importance globally. This study has shown that the SALTAL, along with other instruments, would

be especially valuable for assessing the impact of professional development and other human resource development strategies on the instructional leadership beliefs and attitudes of principals. While the SALTAL was developed and has been validated in the New Zealand context, only small terminology adaptations would be needed to make it accessible in other jurisdictions. However, users in other contexts may need to consider whether school leaders have as much autonomy and independence as they do in New Zealand or whether the New Zealand emphasis on a child-centered pedagogy within a multicultural society restricts the generalizability of the SALTAL. Nonetheless, we would expect that where the cause of instructional or learning leadership has been taken up, the SALTAL could provide a useful mechanism for establishing willingness to adopt this emphasis.

References

Adams, J.E. Jr and Copland, M.A. (2007), "Principal licensing and leadership for learning: the need for coherent policy", *Leadership and Policy in Schools*, Vol. 6 No. 2, pp. 153-95.

Ajzen, I. (2005), *Attitudes, Personality and Behavior*, Open University Press, New York, NY.

Alig-Mielcarek, J.M. and Hoy, W.K. (2005), "Instructional leadership: its nature, meaning, and influence", in Miskel, C.G. and Hoy, W.K. (Eds), *Educational Leadership and Reform*, Information Age Publishing, Greenwich, CT, pp. 29-51.

Alsaker, F.D. (1989), "School achievement, perceived academic competence and global self-esteem", *School Psychology International*, Vol. 10 No. 2, pp. 147-58.

American Educational Research Association (AERA), American Psychological Association (APA) and National Council For Measurement In Education (NCME) (1999), *Standards for Educational and Psychological Testing*, American Educational Research Association, Washington, DC.

Barnett, J.E. and Hixon, J.E. (1997), "Effects of grade level and subject on student test score predictions", *Journal of Educational Research*, Vol. 90 No. 3, pp. 170-74.

Blase, J. and Blase, J. (2000), "Effective instructional leadership: teachers' perspectives on how principals promote teaching and learning in schools", *Journal of Educational Administration*, Vol. 38 No. 2, pp. 130-41.

Bollen, K.A. and Curran, P.J. (2006), *Latent Curve Models: A Structural Equation Perspective*, Wiley-Interscience, Hoboken, NJ.

Butler, R. (1990), "The effects of mastery and competitive conditions on self-assessment at different ages", *Child Development*, Vol. 61 No. 1, pp. 201-10.

Byrne, B.M., Shavelson, R.J. and Muthen, B. (1989), "Testing for the equivalence of factor covariance and mean structures – the issue of partial measurement invariance", *Psychological Bulletin*, Vol. 105 No. 3, pp. 456-66.

Council of Chief State School Officers (CCSSO) (2008), *Educational Leadership Policy Standards: ISLLC 2008*, CCSSO, Washington, DC.

Chen, F., Bollen, K.A., Paxton, P., Curran, P.J. and Kirby, J.B. (2001), "Improper solutions in structural equation models: causes, consequences, and strategies", *Sociological Methods and Research*, Vol. 29 No. 4, pp. 468-508.

Cheung, G.W. and Rensvold, R.B. (2002), "Evaluating goodness-of-fit indexes for testing measurement invariance", *Structural Equation Modeling*, Vol. 9 No. 2, pp. 233-55.

Costello, A.B. and Osborne, J W. (2005), "Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis", *Practical Assessment Research & Evaluation*, Vol. 10, available at: www.pareonline.net/pdf/v10n7.pdf (accessed February 9, 2012).

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum likelihood estimation from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society, Series B*, Vol. 39 No. 1, pp. 1-38.

Donaldson, S.I. and Grant-Vallone, E.J. (2002), "Understanding self-report bias in organizational behavior research", *Journal of Business and Psychology*, Vol. 17 No. 2, pp. 245-60.

Educational Testing Services (ETS) (2009), *Understanding Your School Leadership Series Scores*, Educational Testing Service, Princeton, NJ.

Elmore, R.F. (2004), *School Reform from the Inside Out: Policy, Practice, and Performance*, Harvard Education Press, Cambridge, MA.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.W. (2004), *Applied Longitudinal Analysis*, Wiley, Hoboken, NJ.

Fleenor, J.W., Smither, J.W., Atwater, L.E., Braddy, P.W. and Sturm, R.E. (2010), "Self – other rating agreement in leadership: a review", *The Leadership Quarterly*, Vol. 21 No. 6, pp. 1005-34.

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

Hallinger, P. (2011a), "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*, Vol. 47 No. 2, pp. 271-306.

Hallinger, P. (2011b), "Leadership for learning: lessons from 40 years of empirical research", *Journal of Educational Administration*, Vol. 49 No. 2, pp. 125-42.

Harris, L.R. and Brown, G.T.L. (2010), "Mixing interview and questionnaire methods: practical problems in aligning data", *Practical Assessment Research & Evaluation*, Vol. 15, available at: http://pareonline.net/pdf/v15n1.pdf (accessed February 9, 2012).

Interstate School Leaders Licensure Consortium (ISLLC) (1996), *Standards for School Leaders*, CCSSO, Washington, DC.

Kasanen, K., Räty, H. and Eklund, A.L. (2009), "Elementary school pupils' evaluations of the malleability of their academic abilities", *Educational Research*, Vol. 51 No. 1, pp. 27-38.

Kenny, D.A. and West, T.V. (2010), "Similarity and agreement in self-and other perception: a meta-analysis", *Personality and Social Psychology Review*, Vol. 14 No. 2, pp. 196-213.

Klockars, A.J. and Yamagishi, M. (1988), "The influence of labels and positions in rating scales", *Journal of Educational Measurement*, Vol. 25 No. 2, pp. 85-96.

Lam, T.C.M. and Klockars, A.J. (1982), "Anchor point effects on the equivalence of questionnaire items", *Journal of Educational Measurement*, Vol. 19 No. 4, pp. 317-22.

Leithwood, K., Seashore Louis, K., Anderson, S. and Wahlstrom, K. (2004), *How Leadership Influences Student Learning*, Ontario Institute for Studies in Education, Center for Applied Research and Educational Improvement, Toronto.

Little, T.D., Cunningham, W.A., Shahar, G. and Widaman, K.F. (2002), "To parcel or not to parcel: exploring the question, weighing the merits", *Structural Equation Modeling*, Vol. 9 No. 2, pp. 151-73.

McArdle, J.J. (2007), "Five steps in the structural factor analysis of longitudinal data", in Cudeck, R. and MacCallum, R.C. (Eds), *Factor Analysis at 100: Historical Developments and Future Directions*, LEA, Mahwah, NJ, pp. 99-130.

Marsh, H.W. and Hocevar, D. (1985), "Application of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups", *Psychological Bulletin*, Vol. 97 No. 3, pp. 562-82.

Ministry of Education (2008), *Kiwi Leadership for Principals: Principals as Educational Leaders*, Ministry of Education, Wellington, NZ.

Mitman, A.L. and Lash, A.A. (1988), "Students' perceptions of their academic standing and classroom behaviour", *Elementary School Journal*, Vol. 89 No. 1, pp. 55-68.

Murphy, J. (2003), *Reculturing Educational Leadership: the ISSLC Standards Ten Years Out*, Vanderbilt University, Nashville, TN.

Murphy, J. (2005), "Unpacking the foundations of ISLLC standards and addressing concerns in the academic community", *Educational Administration Quarterly*, Vol. 41 No. 1, pp. 154-91.

Nelson, B.S. and Sassi, A. (2005), *The Effective Principal: Instructional Leadership for High Quality Learning*, Teachers College Press, New York, NY.

Pont, B., Nusche, D. and Moorman, H. (Eds) (2008), *Improving School Leadership* (Vol. 1: Policy and Practice), OECD, Paris.

Quinn, D.M. (2002), "The impact of principal leadership behaviors on instructional practice and student engagement", *Journal of Educational Administration*, Vol. 40 No. 5, pp. 447-67.

Richardson, V. and Placier, P. (2001), "Teacher change", in Richardson, V. (Ed.), *Handbook of Research on Teaching*, American Educational Research Association, Washington, DC, pp. 905-47.

Robinson, V.M.J. (2007), "William walker oration: school leadership and student outcomes: identifying what works and why", in Gurr, D. (Ed.), *National Conference of the Australian Council for Educational Leaders*, Australian Council for Educational Leaders, Melbourne.

Robinson, V.M.J., Eddy, D. and Irving, S.E. (2006), "Catering for diversity in a principal induction programme", *School Leadership & Management*, Vol. 26 No 2, pp. 149-67.

Robinson, V.M.J., Lloyd, C.A. and Rowe, K. (2008a), "The impact of leadership on student outcomes: an analysis of the differential effects of leadership types", *Educational Administration Quarterly*, Vol. 44 No. 5, pp. 635-74.

Robinson, V.M.J., Irving, S.E., Eddy, D. and Le Fevre, D.M. (2008b), "Capability in the leadership of teaching and learning in New Zealand: the validity and utility of a self-assessment tool", in Brundrett, M. and Crawford, M. (Eds), *Developing School Leaders: An international Perspective*, Routledge, Oxon, pp. 155-72.

Singer, J.D. and Willett, J.B. (2003), *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, New York, NY.

Spillane, J., Hallett, T. and Diamond, J. (2003), "Forms of capital and the construction of leadership: instructional leadership in urban elementary schools", *Sociology of Education*, Vol. 76 No. 1, pp. 1-17.

Sung, Y.-T., Chang, K.-E., Chang, T.-H. and Yu, W.-C. (2010), "How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments", *Journal of Adolescence*, Vol. 33 No. 1, pp. 135-45.

Vandenberg, R.J. and Lance, C.E. (2000), "A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research", *Organizational Research Methods*, Vol. 3 No. 1, pp. 4-70.

Wilson, J. and Wright, C.R. (1993), "The predictive validity of student self-evaluations, teachers' assessments, and grades for performance on the verbal reasoning and numerical ability scales of the differential aptitude test for a sample of secondary school students attending rural appalachia schools", *Educational and Psychological Measurement*, Vol. 53 No. 1, pp. 259-70.

Wu, A.D., Li, Z. and Zumbo, B.D. (2007), "Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: a demonstration with TIMSS data", *Practical Assessment, Research & Evaluation*, Vol. 12, available at: http://pareonline.net/getvn.asp?v=12&n=3 (accessed February 9, 2012).

Appendix

No.	Dimension and items
Dimension 1: knowledge and skills for leading teaching and learning	
1	Have a sound and up-to date knowledge of effective teaching and learning
2	Possess a thorough understanding of the New Zealand curriculum framework
3	Demonstrate a good understanding of key concepts used in the current assessment debates
4	Use on-going school-wide assessment to improve teaching and learning
5	Lead information technology (IT) developments in ways that enhance teaching and learning
6	See the development of a school culture focused on learning as a critical factor in creating an effective school
Dimension 2: commitment to ensuring positive learning outcomes for all students	
7	Value the whole student, and use their cultural background to promote engagement with the curriculum
8	Believe that, for all students, learning can be positively enhanced through the principal's influence on the quality of teaching
9	Demonstrate a strong sense of personal responsibility and accountability for the learning outcomes of all students
10	Create opportunities for staff to innovate and experiment with strategies to enhance student learning
Dimension 3: collaborative leadership	
11	Provide or ensure, feedback to teaching staff on teaching effectiveness and student learning
12	Work effectively with the Board of Trustees (BOT) to develop and achieve important school-wide goals
13	Welcome feedback and challenge
14	Align school and local community objectives and cultures to support positive outcomes for students
15	Allocate resources, including funds and time, to enhance effective teaching
16	Ensure parents and caregivers are well informed about the school and the ways they can support student learning processes
17	Facilitate the creation of a collaborative and ambitious vision for the school that is shared by students, staff, parents, the BOT and the community
18	Plan and adopt a key set of strategies to ensure the ongoing professional development of the staff
19	Develop and maintain systems to support the effective operation of the school, based on good management practice and in compliance with all statutory reporting requirements
20	Facilitate change by using sound problem solving skills
Dimension 4: ethical leadership	
21	Lead with integrity
22	Effectively manage your own workload
23	Make and explain the reasons for difficult decisions
24	Hold others accountable, where appropriate

Table AI.
SALTAL items by dimension

About the authors

Gavin T.L. Brown, PhD, is a Psychometric Researcher specialising in educational assessments and affective instrument development. His special interest is in the social psychology of teacher and student responses to assessment. Gavin T.L. Brown is the corresponding author and can be contacted at: gt.brown@auckland.ac.nz

Constance Chai, PhD, was a Research Associate with the First-time Principals Programme, Centre for Educational Leadership, Faculty of Education at the University of Auckland, following her Research Fellowship at the Educational Testing Services, Princeton, New Jersey.