Measuring leadership and management and their linkages with literacy in rural and township primary schools in South Africa

GABRIELLE WILLS
SERVAAS VAN DER BERG

Stellenbosch Economic Working Papers: WP21/2018

www.ekon.sun.ac.za/wpapers/2018/wp212018

December 2018

KEYWORDS: education, management, school quality, leadership, South Africa, literacy
JEL: I21, J24, M12

ReSEP (Research on Socio-Economic Policy)
http://resep.sun.ac.za

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH

SOUTH AFRICA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

www.ekon.sun.ac.za/wpapers

Measuring leadership and management and their linkages with literacy in rural and township primary schools in South Africa[1]

Gabrielle Wills[2] and Servaas van der Berg[3]

December 2018

Abstract

This paper describes a rigorous process to develop and trial new metrics for measuring and codifying school leadership and management practices and processes that are considered theoretically related to literacy outcomes. The predictive validity of these measures is assessed in challenging contexts including 60 township and rural primary schools in South Africa. We observe a randomness to how better leadership and management practices are distributed across better and worse performing schools. Regression analyses confirm weak and inconsistent linkages between measured leadership and management dimensions and literacy outcomes across the sample. However, we find evidence of stronger linkages with intermediate outcomes, including evidence of curriculum coverage. This research contributes to a burgeoning, yet underdeveloped literature on educational management and leadership in Africa and the challenges of measurement in this context.

[1] **Acknowledgements and funding:**

We are grateful to the quantitative and qualitative fieldwork teams who made the larger "Leadership for Literacy" study possible including Nick Taylor, Ursula Hoadley, Jaamia Galant, Francine de Clercq, Nic Spaull, David Carel, and Nompumelelo Mohohlwane. We acknowledge David Carel in particular for managing the quantitative data collection process and Lillie Pretorius in contributing to test development. Thank you to Marie-Louise Schreve, Carine Brunsdon and Silke Rothkegl-Van Velden for their administrative support. We also acknowledge the host of fieldworkers, data capturers and test markers who made this project possible.

This research is funded by the Economic and Social Research Council [grant ES/N01023X/1].

[2] Gabrielle Wills is a researcher with Research on Socio-Economic Policy (ReSEP), Stellenbosch University. Postal address: Department of Economics, Schumann Building, De Beer Street, Stellenbosch University, Stellenbosch 7600. Email: gabriellewills@gmail.com

[3] Servaas van der Berg is a professor in the Department of Economics, Stellenbosch University and leads ReSEP. Email: svdb@sun.ac.za

I. Introduction

Since the 1960s, there has been significant growth in the global knowledge base on school leadership and management. However there is an increasing acknowledgement of the limits of this knowledge, with 'far less systematic knowledge on how school leaders carry out their roles in developing nations throughout the world' (Hallinger, 2017, p. 363). With respect to Africa, Hallinger's (2017) recent review on educational management and leadership identifies the literature as 'emergent', with contributions dominated by one or two countries. But even South Africa, as the largest country contributor to the African literature, derives its knowledge base predominately from studies relying on qualitative research methods, with study locations limited to a few schools (Bush and Glover, 2016). There is still limited understanding of the empirical linkages between school leadership and management (SLM) and learning outcomes in the country and Africa in general (Hoadley et al., 2009), where statistical modelling is seldom used to explore leadership and management effects (Hallinger, 2017).

However, there has been a long-held view in the South African school effectiveness literature that much of the unexplained variation in learner performance across schools, particularly historically disadvantaged schools, may be attributed to differences in unobserved management and leadership competencies (Crouch and Mabogoane, 2001, 1998). Subsequent to this finding (and spurred by the increased availability of large-scale schooling datasets) school effectiveness studies began to include proxies or 'emergent' indicators of SLM practices or processes into education production function models (Gustafsson, 2007; Shepherd, 2011; Spaull, 2013a; Taylor and Prinsloo, 2005; Taylor, 2011; Van der Berg and Louw, 2006). The collective findings from these studies suggest that higher learning levels may be associated with protecting and extending learning time, which includes managing teacher absenteeism and late-coming (Gustafsson, 2007; Van der Berg and Louw, 2006); monitoring and support for planning and delivery in relation to curriculum coverage (Taylor and Prinsloo, 2005); the procurement and management of books and stationary (Kotze, 2017; Shepherd, 2016; Spaull, 2013b; Van der Berg, 2008); and the quality assurance of tests and the monitoring of results. The most prominent positive linkage identified between indicators of SLM and learning relate to increased management of time-on-task, opportunity to learn and monitoring curriculum coverage (Carnoy et al., 2015; Taylor, 2011).

These are useful insights, but the inclusion of these indicator variables in education production function models has been far from sufficient in accounting for large remaining unexplained variation in learning initially attributed to management. Furthermore, what is often not recognised is how many SLM indicators that one would assume to matter for learning are insignificant and excluded from general models (Gustafsson, 2005; Van Staden and Howie, 2014). As Gustafsson (2005, p. 20) reflects, this does not necessarily indicate that these factors are not important but is more about the 'tenuousness' of their measurement, particularly a strong reliance on unreliable principal self-reports of SLM practices and behaviours. An obvious question that arises is whether more rigorous efforts to quantify leadership and management practices, and centring this measurement on clear conceptual frameworks, would aid the detection of stronger linkages between SLM and learning than has previously been identified?

This paper describes a rigorous process to develop and trial new metrics for measuring and codifying SLM practices and processes and their linkages with literacy outcomes in challenging contexts. The first research objective is to measure leadership and management practices and competencies as conceptualised in our 'leadership for literacy' framework (Hoadley, 2018). Through a review of the literature on 'leading for reading', we identify six categories of resources available to school leaders in shaping a literacy learning environment. The development of a rubric provides the critical link

between the conceptual framework and its empirical application as it articulates competencies in allocating, using, and deploying these six types of resources to promote literacy development in the school. This in turn guides instrument development and a process to codify composite indices of 'leadership for literacy' practices from collected survey data.

The second objective is to assess the reliability and validity of derived 'leadership for literacy' indices. Specifically, we identify the predictive validity of the indices using multivariate regression modelling. The data for this process are gathered from a study sample of 60 township and rural public schools accommodating poor students from three South African provinces. In addition to collecting data to code SLM indices, we obtain value-added reading and literacy scores for 631 grade 3 and 2 656 grade 6 students in these schools to establish whether the indices are predictive of literacy levels and gains.

To our knowledge this is the first study to attempt to generate composite quantitative measures of SLM in South African schools that can be used in statistical models. The paper contributes to a knowledge base on leadership and management in challenging primary school contexts in Africa. The findings highlight that despite efforts to measure SLM more rigorously, including composite indices in statistical models rather than just using 'proxy' variables, the linkages with literacy are found to be weak and inconsistent across poor primary school contexts. This result is supported by qualitative case study work in selected study schools. Stronger linkages are found between our SLM indices and intermediate outcomes.

The following section provides background on measuring linkages between leadership and learning, briefly considering three different approaches to instrument development. This is followed by a discussion of methodology and data in section III. The results in Section IV reports information on the internal consistency, reliability and validity of our measurement approach. A specific focus is then given to establishing the predictive validity of the measures. We conclude with a discussion in section V.

II. Background

Qualitative approaches to investigating linkages between leadership and learning yield support for the educational value of leadership, particularly when framed from an instructional leadership perspective (Leithwood et al., 2004; Robinson et al., 2008). Yet studies using larger-scale quantitative data designs often contradict the heroic value placed on leadership and management and its ability to generate student achievement. For example, a meta-analysis of 37 multi-national studies by Witziers et al (2003) found an average leadership effect on student outcomes in the form of a z-score of only 0.02, which reflects no or very weak impact.

The use of narrow theoretical frameworks in conceptualising SLM is one of various explanations provided for the predominance of null-results or weak effects. While theory is at the core of quality empirical study, too little attention is given to research methodology and assessment tools as the critical link to the development of a valid knowledge base underlying the practice of educational leadership and management (Heck and Hallinger, 2005, p. 232).

The inadequacy of existing instruments to measure leadership and management effects has been exposed in the past decade as econometricians begin to isolate out 'principal effects' or programmes identifying causal impacts of SLM interventions on learning. Using very large administrative datasets, principal quality is shown to vary across schools and this quality matters considerably for learning (Branch et al., 2012; Coelli and Green, 2012; Grissom et al., 2015a). In the United States, a higher quality principal can raise student outcomes by about 2 to 7 months of learning in a year. Lower quality principals can lower student outcomes by the same amount (Branch et al., 2013). Recent evaluations

of principal training or management programmes in both developed and developing country contexts also imply that what leaders and managers do, and how they do it, can have significant implications for learning in school environments (Fryer, 2017; Tavares, 2015).

These results imply that instruments developed to capture leadership and management competencies and practice have been limited in the extent to which they can detect the educational value of SLM, although limited sample size and limited variation in practice may also compromise the identification of effects (Robinson et al., 2008), regardless of how valid or reliable an instrument may be.[4]

Beyond the inclusion of 'indicator proxy variables', South African studies have typically not attempted to develop approaches to quantitatively measure school leadership and management in our school contexts. One exception is a qualitative study by Taylor et al (2013). They develop a theoretical framework drawing on an instructional leadership framework by Robinson et al. (2008) to construct a quasi-quantitative measure for management and leadership competencies in 5 matched pairs of case study schools. Their findings pointed to a lack of systematic differences in SLM competencies across better and lower performing schools.

Failure to develop and use instruments to measure SLM as a fuller construct or concept beyond 'indicator variables' in the local context, has arguably hampered the development of policy-relevant SLM research in South Africa. It has not been possible to determine whether the weak explanatory power of SLM indicators reflects the 'tenuousness' of measurement or a lack of variation in management practices and/or learning to identify significant linkages, particularly in historically disadvantaged schools.

In response, the next discussion considers common approaches used in constructing measures of SLM, particularly highlighting a rubric development approach which is adopted in this paper.

Three approaches to measure SLM

Methods to measure or quantify leadership and management in schools are typically designed to reflect the underlying frameworks used to conceptualise SLM constructs. Although conceptual frameworks may vary, limiting comparison across studies, they may still employ common approaches to collect SLM data (Heck and Hallinger, 2005). I consider broadly three different types of methods that have been used.

Self-administered instruments to detect principal effectiveness or principal leadership

Empirical work in educational management and leadership has long been acquainted with the use of self-administered questionnaires, including items to evaluate leadership behaviours and effectiveness (for a list of available instruments see Condon and Clifford (2012)).[5] Although many of these instruments are used in evaluations of principal performance, they may vary in terms of their reliability and validity (Goldring et al., 2009) and even the most rigorously developed instruments have their shortcomings.[6] Ratings tend to be inflated, even when provided by supervisors or peers, so that less

[4] The very large datasets used in principal effect studies allow for the detection of SLM quality variation across populations of schools, rather than smaller school samples typically available to researchers administering SLM instruments in schools.

[5] A closely-related set of instruments asks teachers and supervisors to rate various dimensions of the school climate as outcome indicators of principal performance.

[6] The Vanderbilt Assessment of Leadership in Education (VAL-ED) principal evaluation tool has been documented as one of the most valid and reliable scales in the US context. We explored the possibility of using this scale for our project but this is a proprietary tool with cost implications to use and presenting potential limitations for adaptability and translation in the South African context.

than desirable variation is detected in leadership scores across schools (Grissom et al., 2015b, p. 22; McCullough et al., 2016). However, a review of decades of studies applying the Principal Instructional Management Rating scale in various national contexts suggests that the data collected from teachers typically yields more valid results compared with principal self-reports or supervisor reports (Hallinger, 2008, p. 16).

There is evidence that many of these instruments provide measures that are shown to be associated with learning, although the strength of association may vary depending on the learning outcome measure in question. Positive associations between principal rating scores and learning are more likely to be muted when the outcome in question is value-added in reading and writing rather than mathematics or science; and associations appear to be less likely to be found in primary than secondary schools (Grissom et al., 2015b; McCullough et al., 2016). This calls for learning-centred leadership theories where leadership is conceptualised not only in terms of learning in general as its 'end' but with a subject-specific end in mind (Stein and Nelson, 2003).

Instruments to measure principals' time allocations to different activities

A significant number of studies have attempted to quantify the work of principals by considering their allocations of time to different tasks measured through work activity analysis, structured observation and reflective interviews or self-reported activity logs (Camburn et al., 2010; Grissom et al., 2015b; Horng et al., 2009; Lee and Hallinger, 2012). Time given to a task does not necessarily signal the quality with which those tasks are executed; nevertheless, these studies provide useful insights into leadership and management where two important observations are noted.

First, these studies highlight that SLM tasks comprise a diverse and large set of functions, but there is evidence that higher allocation of time to some tasks over others may yield returns in terms of improved learning (Grissom et al., 2015b). Second, principal time use and allocations are shown to vary substantially across societies and these patterns of behaviour are influenced by economic, socio-cultural and the institutional features of their societies (Lee and Hallinger, 2012). Although internationally standardised approaches to measuring SLM are obviously very useful for comparative purposes, developing contextually relevant approaches and instruments to measuring these constructs certainly has merit. The third useful insight is that leadership and management tasks vary notably over time within the same principals (Camburn et al., 2010).

Descriptive scoring rubrics to assess SLM

Scoring rubrics are increasingly being used internationally to quantify competencies in areas of education management, assessment, or other systems technologies (Arcia et al., 2015; Bloom and Van Reenen, 2010; Lemos and Scur, 2017). A key benefit of a rubric is that many sources of data can be combined to assess how an institution or practices compare to a described benchmark where the rubric description guides the data to be collected. This contrasts with other measurement approaches that often depend on one source of data such as teacher perceptions of their manager or principal self-reports to provide metrics of quality or competence.

The World Bank's SABER (Systems Approach for Better Education Results) program, for example, collects comparable data on the policies and institutions of education systems around the world and benchmarks them against good practice using a rubric scoring approach (Arcia et al., 2015; Arcia and Demas, 2013). The World Management Survey (WMS) collects data on the level of formalisation of management processes across private and public-sector industries relying on a rubric to score institutions on their management practices and processes (Bloom et al., 2014). This measurement approach has provided convincing evidence to date that school management practices can be

measured meaningfully and are linked to learning outcomes and better teacher practices, even in developing country[7] contexts (Uganda, Brazil and India) (Bloom et al., 2015; Crawfurd, 2017; Scur, 2017).

The success of this study likely depends on its reliance on highly competent researchers as interviewers, resulting in reliable score measures. It involves a telephonic semi-structured interview process with a school principal/head teacher which requires multiple high-level judgements from a highly competent interview researcher to score a school's management practices using a pre-determined rubric. However, in challenging contexts with compromised telephonic infrastructure and a limited supply of highly skilled researchers or fieldworkers, the opportunity for telephonic calls with principals may be limited. Hiring high-level researchers to conduct fieldwork may also require research budgets that may not be typically available in poorer country contexts.

Our approach to measuring SLM is operationalised in a manner conducive to reducing the cognitive load imposed on fieldworkers during data collection.[8] We use fieldworkers to administer face-to-face interviews and collect data. Compensating for the use of fieldworkers rather than high-level researchers we aimed to add reliability to the scoring approach through combining data from various respondents through interviews or self-administered questionnaires and collecting 'evidence' of practice rather than just relying on self-reports from principal interviews. In this respect our rubric approach draws on two of the three mentioned methods to generate SLM measures.

III. Method and Data

SLM measurement approach

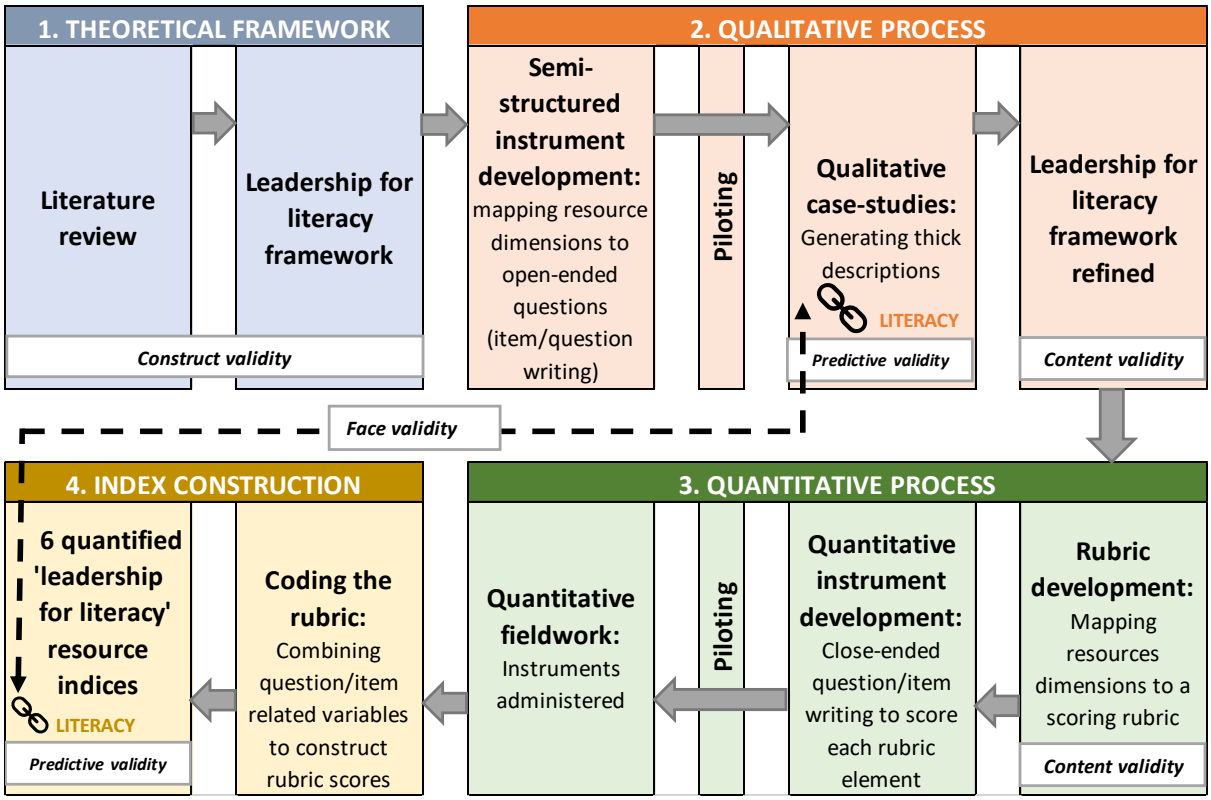
Our SLM measurement approach was operationalized in four phases as shown in Figure 1, including the development of a theoretical framework, its empirical application across interconnected qualitative and quantitative fieldwork components, and final index construction.

[7] Lemos and Scur (2017) point out the additional challenges of measuring management in less functional education systems when poor management practices tend to dominate:

> *"Although the global context of the WMS project allows for a very useful comparison of world-class and poorly managed organizations across a number of countries, the very thick (almost truncated) left tail for developing countries makes it harder to explore the variation of managerial practices in the less well managed organizations"* (Lemos and Scur, 2017, p. 3).

[8] The limited supply of high-quality fieldworkers for large-scale data collection processes in schools is a major constraint to obtaining high-quality school data in South Africa. Using Masters or PhD students for fieldwork is problematic as school fieldwork periods generally coincide with examination or thesis submission times at universities. Other potential fieldwork candidates with educational backgrounds and experience are often full-time employed in the education sector. With 11 official languages, potential fieldwork candidates who are available often do not share the same language proficiencies as those of teachers or students in the sampled school group.

Figure 1: Four phases of a mixed-methods study to develop indices of leadership and management



Phase 1 – Theoretical framework:

Our measurement approach was strongly underpinned by a 'Leadership for literacy' (henceforth LL) theoretical framework, developed through a thorough review of educational studies that explore the processes or resources available to school leaders in promoting reading and literacy across a school (Hoadley, 2018). The literacy focus of this framework draws on Stein and Nelson's (2003) notion that leadership effectiveness is strongly tied to knowledge resources that are subject specific – a term they have identified as 'leadership content knowledge'. As they reflect

> 'Without knowledge that connects subject matter, learning and teaching to acts of leadership, leadership floats disconnected from the very processes it is designed to govern' (Stein and Nelson, 2003, p. 446).

The conceptual framework identifies four kinds of resources available to school leaders in promoting literacy in the school:

1) Material resources:
    a. The allocation of sufficient time for teaching language and reading, as well as protecting and maximising the use of time for learning in general.
    b. The prioritisation and procurement of text resources to support a programme of reading.
    c. The optimal utilization of text resources for reading.
2) Knowledge resources: Own expertise in understanding the value and technology of teaching reading and the extent to which this understanding is shared by school leaders and with teachers across the school.

3) Human resources: The effective recruitment, utilisation and development of teaching expertise, particularly in reading and language.
4) Strategic resources: The mobilisation of available resources to drive a coherent literacy programme.

The theoretical framework implicitly assumes that leadership and management functions are shared across individuals in the school organisation. Expert leadership in teaching reading, for example, may extend beyond manager job-titles to teachers or support staff. In this respect, the measurement approach explores whether leadership for literacy competencies are present within the organisation rather than tied to a specific leader such as the school principal.[9] Our approach therefore cannot be used as principal evaluation tool.

The literacy focus is also contextually appropriate given the extremely low levels of basic literacy that characterise the South African schooling system (Howie et al., 2017). Exploring whether effective leadership and management processes can improve literacy trajectories is an important contribution to the education debate.

It is acknowledged at the outset that management and leadership are referred to interchangeably in our measurement approach. These are distinct concepts, as Hoadley and Ward (2009, p. 9) explain, where "leadership can be exercised throughout the school, by different people at different levels, while management, in contrast, is a structural position, which carries with it specific roles and responsibilities." We have measured outcomes and processes associated with both leadership and management activities that are identified as necessary to supporting a literacy learning environment in the theoretical framework.

Phase 2 – Qualitative process:

The next step in the process involved the development of semi-structured interview and observational instruments to explore leadership for literacy practices in schools, as expressed in the theoretical framework. Question-writing involved the joint work of four South African-based education specialists who also directly engaged in piloting the instruments in township schools. The final set of instruments were administered in four pairs of high and lower achieving case-study schools which were selected from the larger 60 school sample.[10] Thick descriptions were generated through these case studies, adding more specificity to the four framework areas.

Phase 3 – Quantitative process:

The rubric development process involved mapping each of the theoretical resource dimensions as envisaged in the framework into detailed descriptions of competence or evidence of implementation of good SLM practice. The six dimensions are broken down into 114 areas to be evaluated which we refer to as rubric 'elements'. As an example, Table 1 provides a scoring rubric for five elements under the human resource dimension, showing how descriptions relate to quantitative scores of 1 (low) to 5 (high). The sub-dimensions as described in Table 2 provide more information on what we aimed to evaluate. Guided by the rubric descriptions close-ended questions were written, that would yield data necessary to evaluate if a school should be scored 1, 2, 3, 4 or 5 on each rubric element. The use of

[9] This is contextually relevant where policy envisages school leadership and management functions as distributed across members of a school management team comprising the principal and a set of middle-managers (deputy principal and Heads of Department) (Department of Basic Education (DBE), 2016).
[10] Literacy pre-test scores and measures of student socio-economic status were used to select high and low-performing pairs that were as best matched as possible in terms of SES as well as location as the sample allowed.

predominately close-ended rather than open-ended questions reduces the need for probing and research judgements from fieldworkers.

The necessary set of close-ended questions was allocated across six instruments, designed to be administered during one school day. These include a principal interview instrument, deputy principal instrument, a grade 3 teacher interview of the tested class, grade 6 teacher interview of the tested class, a self-administered survey for all teachers and school observational instruments administered by the fieldworker.

A dedicated fieldworker, trained over a four-day process on the instruments, was responsible for conducting SLM interviews and collecting observation data for each school. The training of fieldworkers was supported through an in-school simulation day. Senior project researchers were present during the school-simulation to monitor the quality of interviews and data collection process.

Phase 4 - Index construction

Once data is collected and cleaned, a coding process is used to combine data collected across various instruments to 'objectively' score each rubric element. The process is objective in the sense that the data determines each school's score for a rubric element rather than a researcher making subjective judgements of competence.

In total, over 500 variables collected across the various instruments were used to code 114 rubric elements which range from 1 (lowest possible score) to 5 (highest possible score). The number of elements evaluated in each framework dimension, and the types of data gathered to score each element which include self-reported[11], observational or evidence-based data or a combination of the two, are shown in Table 3. Almost half of the elements are coded using data that are triangulated in some way; for example, using responses from multiple respondents.

The six leadership for literacy dimensions are obtained using a statistical procedure called principal components analysis (PCA) to weight each index in terms of the variation it explains in an underlying unobserved factor. Due to potential inaccuracies in self-reported data from teachers and school leaders or managers, a second set of indices were also combined using only elements that could be evaluated using 'evidence-based' data. Figure 2 plots the distributions of the six "leadership for literacy" indices using different methods to combine the rubric elements. All indices are normalised (i.e. expressed in z-scores).

[11] Even if questions are administered through an interview format, we still consider responses to be self-reported if the questions require gathering respondents' recall information on their experience or perceptions of SLM processes or practices.

Table 1: Example of the qualitative scoring rubric descriptions for 5 elements associated with the human resource index

	Score 1 (LOW)	Score 2	Score 3 (Middle)	Score 4	Score 5 (HIGH)
The presence of a reading expert - there is an identified expert within the school assisting teachers with their reading instruction.	i) Respondents (0 of 3) identifies one or two specific people by name as being best at teaching reading in the school.	i) Respondents (1 of 3) identifies one or two specific people by name as being best at teaching reading in the school.	i) Respondents (2 of 3) identifies one or two specific people by name as being best at teaching reading in the school.	i) Respondents (2 of 3) identifies one or two specific people by name as best at teaching reading but ii) 1 respondent neither identifies a reading specialist nor identifies 'everyone' as good at teaching reading.	i) Respondents (3 of 3) identifies one or two specific people by name as best at teaching reading
Use of a reading expert – assesses whether reading experts actually provide reading instruction to teachers or learners.	i) NA - No reading expert	i) There is a reading expert but ii) no respondents (0 of 3) indicate that the reading expert helps teachers with how to teach reading 'quite a lot' or 'a lot'. But may indicate that the teacher supports *students* with their reading 'quite a lot' or 'a lot'.	i) There is a reading expert but ii) respondents (at least 1 of 3) indicate that the reading expert helps **teachers** how to teach reading 'quite a lot' or 'a lot'.	i) There is a reading expert but ii) **some but not all** respondents indicate that the reading expert helps teachers with how to teach reading 'quite a lot' or 'a lot'.	i) There is a reading expert and ii) respondents (3 of 3) indicate that the reading expert helps **teachers** with how to teach reading 'quite a lot' or 'a lot'. OR 80% or more of language educators indicate they go to the reading specialist for help at least once a term.
Qualifications - The qualifications, specialisms and training of educators teaching African or English language suggest requisite expertise to teach reading in the school.	Less than a quarter of educator respondents teaching African or English language have either i) completed an Advanced Certificate in Education (ACE) or short course in teaching language or reading OR ii) identify that English or African language	25 - 49% of educator respondents teaching African or English language have either i) completed an ACE or short course in teaching language or reading OR ii) identify that English or African language was one their subject majors	50 -74% of educator respondents teaching African or English language have either i) completed an ACE or short course in teaching language or reading OR ii) identify that English or African language was one their subject majors	75% or more (but not all) educator respondents teaching African or English language have either i) completed an ACE or short course in teaching language or reading OR ii) identify that English or African language was one their subject majors	All educator respondents teaching African or English language have either i) completed an ACE or short course in teaching language or reading OR ii) identify that English or African language was one their subject majors.

	was one their subject majors				
The school has a system for acknowledging its best teachers through rewards/awards.	School has NO system for acknowledging its best teachers through rewards/awards other than usual Integrated Quality Management System (IQMS).	School has a system for acknowledging its best teachers through rewards/awards other than usual IQMS.	School has a system for acknowledging its best teachers through rewards/awards other than the usual IQMS (a weak system imposed evaluative mechanism). Singles out best individuals rather than just awarding all or most teachers.	School has a system for acknowledging its best teachers through rewards/awards other than usual IQMS. Singles out best individuals rather than just awarding all or most teachers. These awards/rewards happen regularly (once or more than once a year).	School has a system for acknowledging its best teachers through rewards/awards other than usual IQMS. Singles out best individuals rather than just awarding all or most teachers. These awards/rewards happen regularly (once or more than once a year). Clear criteria are used to determine who gets an award.
The School Governing Body (SGB) supports good hiring as indicated by the principal.	The SGB does not have necessary competencies and skills to make good recommendations on staff appointments (as identified by principal). SGB does not have necessary competencies and skills to fulfil their functions. The school would be identified as 'much better' or 'a little better off' if the school had No SGB.	The SGB does not have necessary competencies and skills to make good recommendations on staff appointments but may have competencies in other areas.	The SGB has necessary competencies and skills to make good recommendations on staff appointments.	The SGB has necessary competencies and skills to make good recommendations on staff appointments AND the SGB has necessary competencies and skills to fulfil their functions. The school would be identified as 'much worse off' or 'worse off' if the school had no SGB.	SGB has necessary competencies and skills to make good recommendations on staff appointments AND the SGB has necessary competencies and skills to fulfil their functions. AND the school would be identified as 'much worse off' or 'worse off' if the school had no SGB.

Table 2: Description of sub-dimensions across 6 "Leadership for Literacy" dimensions

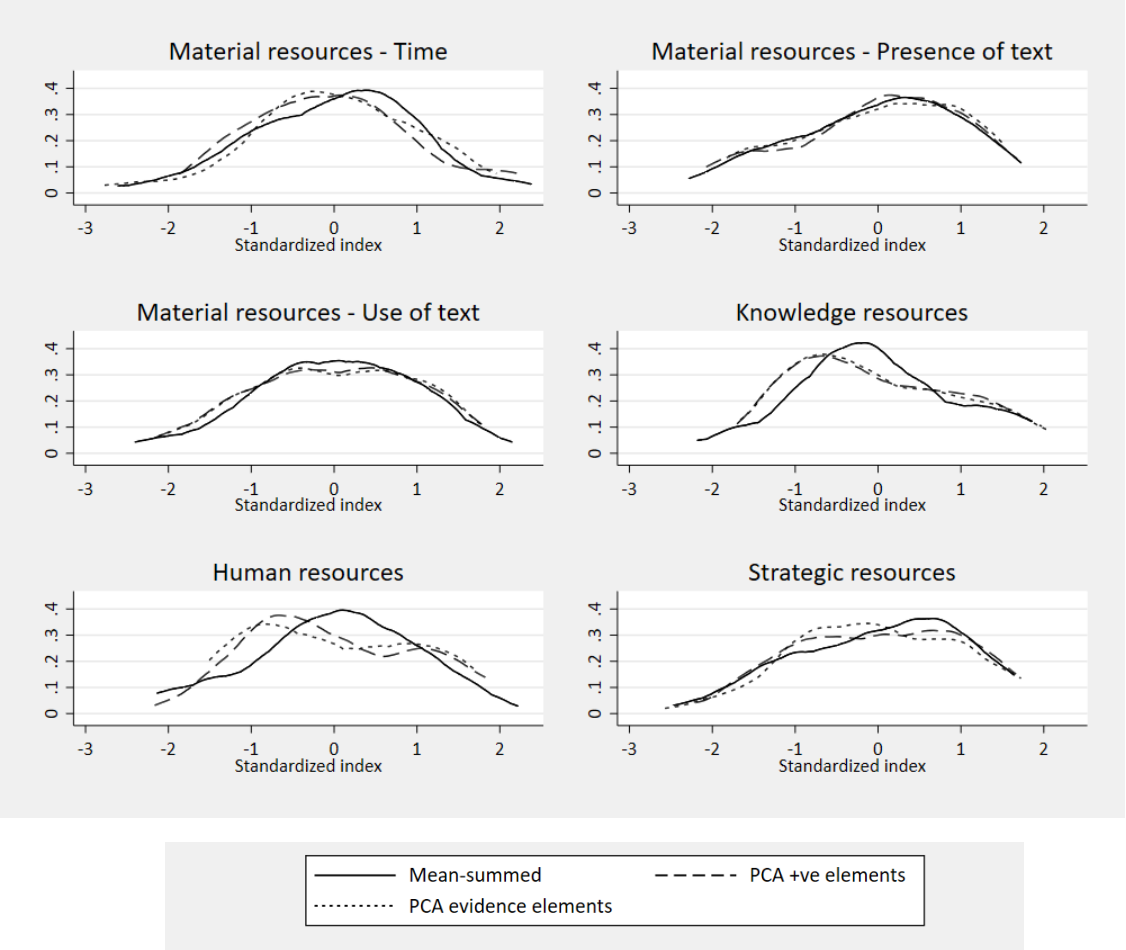
Leadership for Literacy Index Dimensions	Sub-dimension	N elements
Material resources: Time	Allocation / structure of time for teaching of language and reading	5
	Maximum use of teaching time (limited disruptions and few free periods)	4
	Low absenteeism and teacher presence in classroom	5
	Additional time for reading beyond class	5
Material resources: Presence of text	Presence of text in Grade 3 classroom	10
	Presence of text in Grade 6 classroom	9
Material Resources: use of text	Use of text in grade 3 classroom	3
	Use of text in grade 6 classroom	3
	School-wide management of resources to promote availability and use of text	3
Knowledge Resources	Culture of reading among teachers	9
	Knowledge of teaching reading	7
	Knowledge of remediation	2
	Knowledge sharing - professional collaboration	4
Human resources	The presence of managers and leaders in the school to promote reading	4
	Qualifications levels, teacher and school management team (SMT) alignment to subjects and phase specialisations	4
	Presence of reading expert/s in the school	2
	Acknowledging and rewarding teacher performance	2
	Professional development - Teacher exposure to professional development opportunities including workshops on reading instruction	6
	Managing poor performance and consequence management	4
	Appointing staff and attracting talent to promote improved teacher quality	3
Strategic Resources	Use of networks and financial mgt. to support a reading programme.	4
	Evidence of reading assessment practices	3
	Performing tracking of parameters, including reading and reviews of performance (whole staff meetings and one on one meetings).	3
	Monitoring of lessons and curriculum coverage	4
	Clear strategies to create a reading programme (programmes implemented, celebration of reading, promoting enjoyment of reading)	5
	Vision, goal setting and expectations - school goals incorporate 'improved reading'	4

Table 3: Types of data used to construct rubric elements

	N. elements	Data types used to construct rubric elements: N. elements based on respondents' experiential recall	Data types used to construct rubric elements: N. elements based on observational or evidence presented to fieldworker	N. elements triangulated
Material Resources:				
Time	19	18	3	11
Presence of text	19	11	8	1
Use of text	9	13	8	1
Symbolic Resources:	19	3	9	2
Human Resources:	16	16	9	12
Strategic Resources:	32	31	7	23
Total	**114**	**92**	**44**	**50**

Notes: Multiple forms of data can be used to construct elements of the rubric so that row and column totals do not add up to the total number of elements in the rubric. An element is considered 'triangulated' if coding requires using either responses from more than 1 person or different sources of data such as interview and observational data were used in constructing the rubric element.

Figure 2: Distribution of standardised indices (using three different approaches to combine rubric scores)



Notes: The distribution labelled mean-summed, indices are constructed by merely summing scores across all dimension elements and obtained the normalised z-score. 'PCA +elements' shows index distributions when elements are combined using principal components analysis but excluding any elements that load negatively on the first component.

Assessing reliability

We did not administer two rounds of the same instruments across the same schools, so that evaluating the test-retest reliability of the measurement approach – i.e. the degree to which the measurement approach produces stable and consistent results - is not possible. However, we do consider whether the indices discriminate in ways that we would expect, such as observing better practices in rural as opposed to urban schools or in poor schools compared with less poor schools.

Our theoretical framework is multi-dimensional in its conceptualization. Conducting tests of internal consistency, such as Cronbach's alpha which assume construct unidimensionality are not appropriate here (Peters, 2014).

Assessing validity

Methods to establish the validity of the measurement approach were incorporated into each of the four phases in Figure 1. The literature review, its qualitative application in schools and the rubric development process were critical to establishing *content validity*. This is also supported through several rounds of piloting instruments in no-fee schools[12]. The mixed-method design of the study allows for checks of *face validity* – whether our measurement approach appears at face value to measure what it claims to. We also evaluate the *predictive validity* of the SLM measures. We explore how each of the quantified dimension indices and sub-elements are related to literacy outcomes in schools. The method and analytical strategy used to establish predictive validity are now discussed.

School sampling

A possible reason for the identification of weak associations between school management and leadership and student learning is that study samples selected lack variation in both student learning and SLM practices. The 60 schools were purposively selected to artificially add as much student performance variation as possibly exists in the available sampling frame of public schools reaching poorer student populations in three provinces.

We engaged in an intensive search to identify the 30 best possible high-performing no-fee schools in three provinces (KwaZulu-Natal, Gauteng and Limpopo)[13]. This search process (described in detail in Wills (2017)) relied on identifying top performing schools in system-wide low stakes testing data – namely the Annual National Assessments - corroborated against a large dataset we collected of recommended 'good' schools from a host of sources (district officials, school principals and administrative clerks, education related NGOs, unions, other stakeholders, secondary schools performing well in the school-leaving[14] certificate). Then 30 lower performing schools located in similar geographic locations as the higher performing pairs were included in the sample. The schools were also selected based on either their language of instruction or the dominant student language in the school being Zulu, Sepedi or Xitsonga.

[12] There are two groups of public schools in South Africa: no-fee schools that are also synonymous with quintile 1 to 3 schools that are not allowed to charge school fees. They are typically located in poorer areas, accessible to students from lower socio-economic status households. Fee-charging schools (also known as quintile 4 and 5 schools) have historically served white and to a lesser extent Indian and Coloured populations of students. These schools can charge fees but receive a lower per pupil state allocation.

[13] The three provinces are chosen to represent distinct levels of administrative functionality: Gauteng (a highly functional administration), KwaZulu-Natal (medium functionality) and Limpopo (low functionality).

[14] The National Senior Certificate or otherwise known as matriculation examination.

Literacy data

Pre-test literacy and reading scores were obtained at the beginning of the 2017 year in each of the 60 schools. This process was necessary to verify the quality of the selected schools and to select case-study schools for the qualitative work. The availability of post-test scores collected towards the end of the same school year allows us to determine how predictive SLM quantitative measures are not only of literacy levels but of literacy gains.

The grade 6 English literacy test consisted of a silent reading comprehension test and vocabulary test administered to an entire class of grade 6 students in each school. Of the original pre-test sample of 2 656 students, 2 379 wrote the post-test, indicating a low attrition rate of 11%. The two comprehension tests consisted of released items from previous rounds of the grade 4 PIRLS assessment. Permission was received from the IEA for their use. The reliability of these assessments is reflected in a high correlation between pre-test and post-test scores[15].

Additionally, pre- and post-test English Oral Reading Fluency (ORF) scores are available for 599 grade 6 students. ORF tests in African language were also administered to these students but only at the end of the year. At the grade 3 level a battery of tests – including ORF and single word recognition in English and African language - were administered. Pre- and post-test scores are available for 631 grade 3 students.

Analytical strategy

To assess the predictive validity of the leadership for literacy dimensions we use an education production function framework where grade 6 literacy outcomes are expressed as a function of a specific Leadership for Literacy index, LL_{is}, and a set of individual or home, and school characteristics:

$$Y_{ist} = \beta_0 + \beta_1 LL_{is} + \beta_1 I_{is} + \beta_2 S_s + \varepsilon_{is} \quad (1)$$

Here Y_{is} represents the standardized post-test scores[16] of the i_{th} student in school s. I_{is} is a vector of the student's background characteristics including a measure of wealth[17] and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student gets help with homework from someone at home, lives with his/her mother, lives with his/her father, the student's parents' employment status, rural vs. urban location, and whether the student speaks English at home `sometimes' or 'always' compared with 'never'. School controls, S_s include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English.

Equation (1), however, merely estimates *levels* of learning and is unlikely to control for various unobserved factors such as higher learning ability or an historical accumulation of exposure to different teachers or socio-economic factors which may play a stronger role than school processes in determining a student's level of performance. For this reason, a value-added model is also used to estimate whether 'Leadership for literacy' indices and variables of interest explain any differences in literacy skills gained within a school year across the 60-school sample. In this value-added production function framework reflected in equation 2, we include a pre-test score Y_{ist-1}.

$$Y_{ist} = \beta_0 + \beta_1 Y_{ist-1} + \beta_2 LL_{is} + \beta_3 I_{is} + \beta_4 S_{is} + \varepsilon_{is} \quad (2)$$

[15] The Pearson correlation statistic between the pre-test and post-test was almost 0.90.
[16] Post-test scores are standardized in terms of the pre-test score mean and standard deviation.
[17] The wealth index is derived from questions on asset ownership in the student's home constructed using principal components analysis.

Although we assess the presence of a 'direct' relationship between the Leadership for Literacy indices and literacy outcomes, we also explore the possibility of linkages with intermediate outcomes using the following question:

$$Intermediate_s = \beta_0 + \beta_1 LL_s + \beta_2 + \varepsilon_s \quad (3)$$

Here, Y is replaced with 1 of 4 intermediate variables which vary at the school level (not the student level as is the case with equation 1 and 2). School level controls as well as student characteristics averaged to the school level are represented by S_s. The four intermediate variables include:

- Coverage of work in the best grade 6 learner's exercise / workbook (expressed in centiles).
- The percentage of utilised classrooms with a present teacher (and students engaged in a learning activity).
- The percentage of educators in a school indicating that the amount of work they have covered in the curriculum is checked by a head of department (HoD) at least twice a week.[18]
- An index of average teacher 'employee engagement', also referred to as work engagement or worker engagement, augments the traditional notion of "job satisfaction" to consider the connectedness of employees to their organisation and work itself (Harter et al., 2002; Kahn, 1990). An engagement score is generated for teachers in each school and averaged to the school level and normalised. The index is constructed from 8 questions or statements related to work enjoyment, satisfaction with the school's SMT, sense of safety, stress levels, receiving praise or recognition, and feeling valued.[19]

IV. Results

Reliability

One way to assess whether the 'leadership for literacy' indices are reliably measuring what we intend them to is to explore differences in index scores across rural and urban schools, and the poorest 20% of schools against the least poor 20% of schools in the sample. In general, we would expect somewhat higher scores for urban and wealthier schools compared to rural and poorer schools. We explore whether patterns are more aligned with expectations when only evidence-based data is used to

[18] The question asked of educators was "How often does your HoD (i.e head of department) in this school check to see how much of the curriculum you have taught?"

[19] The employee engagement index is derived from the educator questionnaire using the following questions or statements requiring a likert-type response:

- How much do you like teaching in this school?
- When you wake up on a Monday morning, how do you feel about going to work?
- How happy are you with this school's SMT?
- How safe do you feel in this school?
- How stressed do you feel in this school?
- When was the last time you received recognition or praise from school management for doing good work?
- To what extent do you agree with the following statement? "I know exactly what is expected of me from management in this school."
- To what extent do you agree with the following statement? "Our principal makes me feel that I play an important role in this school."

construct elements and in turn the indices (i.e. excluding elements drawing on potentially less reliable self-reported data).

The expected patterns hold for four of six indices, as seen in Table 4. The index 'material resources - presence of text' yields patterns we would expect with significantly more text presence in urban as opposed to rural schools or in wealthier schools. There is evidence of slightly better knowledge resources in urban and wealthier schools. The "material resources – use of text" index favours urban and wealthier schools but these differences are not statistically significant. However, the human resources and strategic resources indices are higher in poorer than wealthier schools – a pattern we would not expect.

Limiting the indices to being constructed from 'evidence-based' elements yields more reliability for 1 of 6 'leadership for literacy' indices, namely strategic resources.

Table 4: Leadership for Literacy indices by location, and school wealth

		Rural	Urban	Poorest 20%	Least poor 20%
Material Resource: Time (z-score)	PCA	-0.03	0.04	-0.02	-0.15
	Evidence-based	-0.04	0.04	-0.27	0.19
Material Resources: Text Presence (z-score)	PCA	-0.5	0.6***	-0.6***	0.5
	Evidence-based	-0.4	0.5***	-0.4**	0.4
Material Resource: Text Use (z-score)	PCA	-0.16	0.19	-0.05	0.43
	Evidence-based	-0.18	0.22	-0.07	0.47
Knowledge Resources (z-score)	PCA	-0.02	0.02	-0.26	0.21
	Evidence-based	-0.10	0.12	-0.33	0.28
Human Resources (z-score)	PCA	0.33	-0.4***	0.4***	-0.53
	Evidence-based	0.36	-0.4***	0.4***	-0.70
Strategic resources (z-score)	PCA	0.43	-0.5***	0.4*	-0.25
	Evidence-based	-0.38	0.5***	-0.4***	0.60
	N schools	33	27	18	13

Notes: Differences are statistically significantly different at *** p<0.01, ** p<0.05, * p<0.1.

Face validity

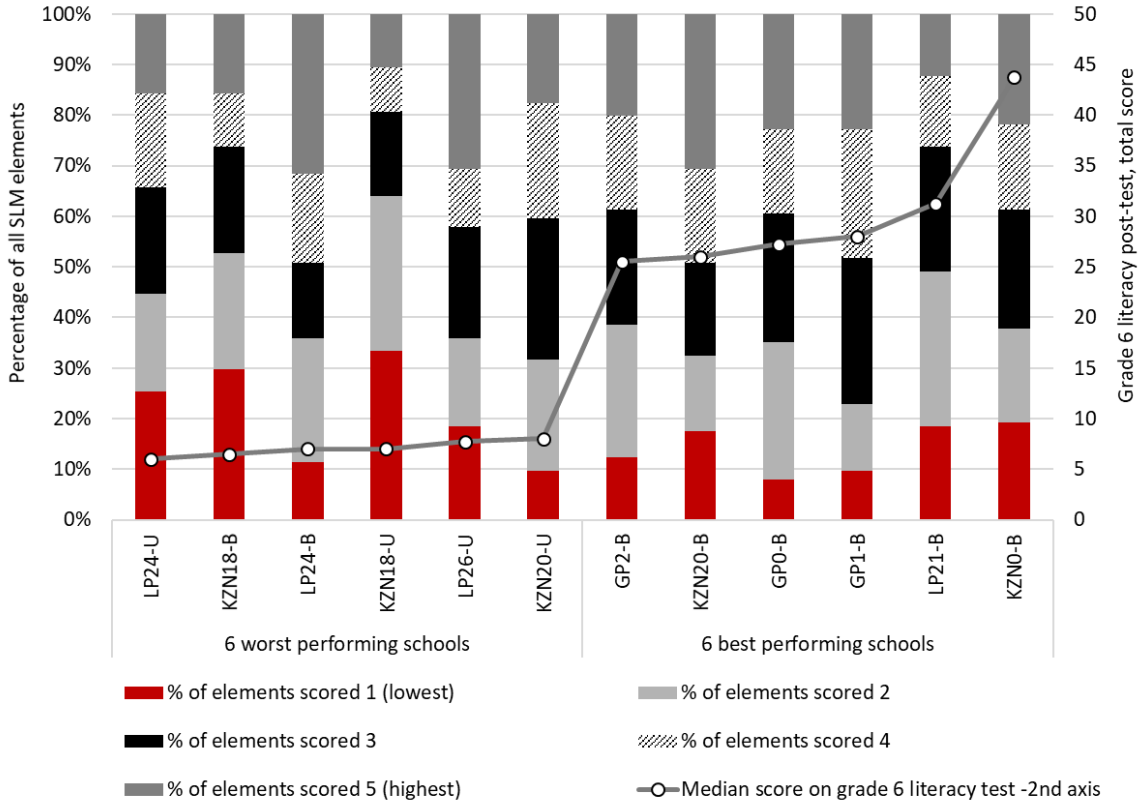
Comparing the leadership for literacy scores for the 8 case-study schools against qualitative findings on these schools is important in establishing the validity of the quantitative SLM measurement approach. When drilling down to each of the framework dimensions, the quantitative measurement results often contradict the findings from the quantitative analysis. In a mixed-methods paper we examine convergences and divergences between the quantitative data collected in instruments and the qualitative findings relating to 8 topics that were probed in the case-study school qualitative interviews (Taylor et al., 2018). We identified convergence between the qualitative and quantitative findings for 2 of the 8 topics compared. The convergence on these two topics was not surprising as they could be assessed using observational data. The other sub-dimensions require collecting self-reported recall, experiential or perception-based information on constructs or topics that cannot be directly observed, opening the door for less reliable *socially acceptable responses*. Given these face-validity concerns, the analysis explores whether the predictive validity of the indices improves if we only include evidence-based elements in the construction of indices (i.e. excluding elements derived from self-reported perception-based or recall responses).

Predictive validity

Descriptive findings

The descriptive results suggest that overall systematically better practices are not detected in better performing schools. There is a randomness in how better practices appear to be distributed across the school sample. This is starkly illustrated in Figure 3, which shows the percentage of all 114 rubric elements scored 1 (lowest), 2, 3, 4 and 5 (highest) for the 6-best and 6-worst performing schools (ranked by the performance of the middle learner in the grade 6 English literacy post-test). The best performing schools are no more likely to a have a larger percentage of the highest possible scores than the 6-worst performing schools.

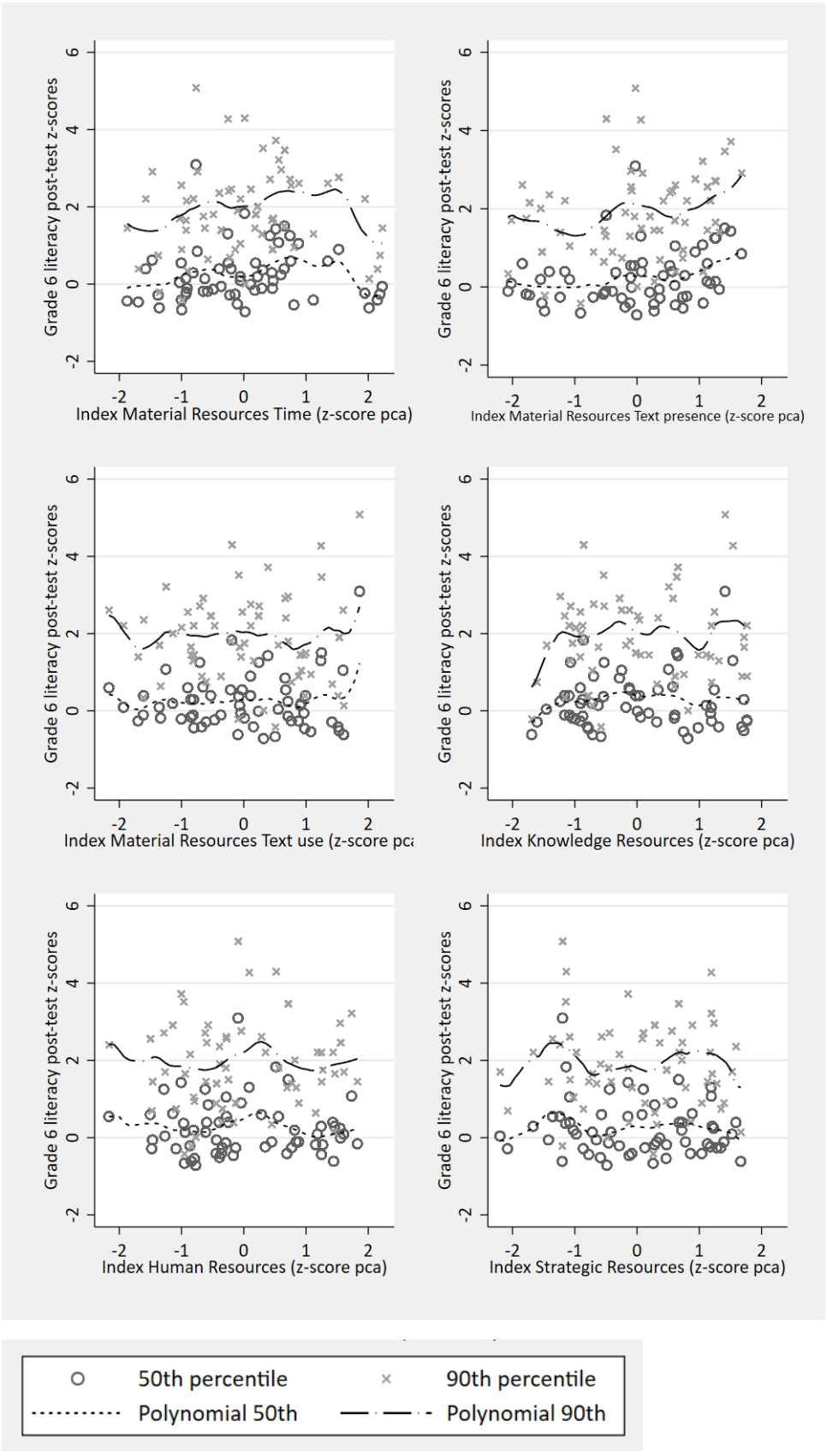
Figure 3: Leadership for literacy scores across 114 rubric elements for the 6-best and 6-worst performing schools



Source: Leadership for literacy dataset. **Notes**: 60 schools are considered in the study, but results are only shown for the 6 best and 6 worst performing schools, ranked by the median performance of grade 6 learners in a reading comprehension and vocabulary test.

There is also little descriptive evidence of a positive association between the six framework dimensions and literacy outcomes. Figure 4 illustrates the raw correlation between the six Leadership for Literacy dimensions and grade 6 English literacy outcomes (expressed as a normalised reading comprehension and vocabulary test score) at 50th and 90th percentile. A local polynomial regression line is fitted separately for the 50th percentile (median) and 90th percentile scores. The lack of any systematic relationship for at least 5 of the 6 dimensions is evident. The presence of text in the school (material resources) is slightly positively associated with grade 6 literacy scores but in the multivariate estimations this association is absorbed by controls for socio-economic status.

Figure 4: Median and 90th percentile grade 6 literacy post-test scores plotted against six Leadership for Literacy indices for 60 schools



Source: Leadership for literacy dataset.

Multivariate estimation of literacy outcomes

The overall impression of no or weak associations between the leadership for literacy indices and learning is confirmed in the multivariate analyses. If positive and significant results do emerge they are typically inconsistent across grades and different tests. This is seen in Tables 5 and 6.

Coefficients on indices for estimates of grade 6 literacy outcomes are shown in Table 5. Grade 6 outcomes include English written literacy test scores (a combined reading comprehension and vocabulary test) expressed in z-scores; English Oral Reading Fluency (ORF) scores expressed in z-scores; and African language ORF test scores expressed as the percentage of words read correctly per minute (% WCPM). Estimates of grade 3 literacy outcomes including English and African language ORF (% of WCPM) are shown in Table 6. In each table there are estimates of literacy levels and value-added estimates. Results are always shown for 2 models. The first model excludes any other controls, highlighting the raw correlation between each index and literacy outcomes. The second specification includes a full set of individual, classroom and school controls. Coefficients are shown on the pre-test score obtained by the student at the beginning of the year on the same test.[20]

In Table 5, the lack of explanatory power of the indices is highlighted by the low R-squared values in model 1. The human resources index shows the most promise with significant and positive coefficients of 0.1 standard deviations in the value-added estimations of grade 6 literacy with full controls. However, no significant coefficients on any of the indices are observed for estimates of grade 6 English and African ORF.

In estimations of grade 3 English and African language ORF, a positive coefficient emerges on the strategic resources index in estimations of ORF proficiency levels. It suggests a 3-4% point improvement in the number of words read correctly per minute for a standard deviation increase in the strategic resources index, but this positive association is not evident in the value-added model. The most consistent positive association is observed in the effective use and allocation of time. A standard deviation increase on the 'material resource - use of time' index is associated with a 1-2% point increase in words read correctly per minute in African language and English ORF. At the grade 3 level, the efficient use and allocation of time, specifically for reading and language, appears not only to matter for ORF proficiency levels but for better reading during a school year. However, the magnitude of the association remains small. Coefficients on other indices are also often negative in magnitude although not statistically significant in value-added models.

We explored whether the influence of the 5 leadership for literacy dimensions on literacy outcomes, may be mediated through the presence of knowledge resources (i.e. including interactions between 5 of the dimensions and the index for knowledge resources). We find no evidence of this (tables not shown). A similar result holds if we replace the knowledge resources index with grade 6 teacher vocabulary test scores.

Our results suggest that the exclusion of self-reported data in constructing indices is unlikely to significantly improve the predictive validity of the measurement approach. Limiting the indices to be based on evidence-based elements only slightly improves the validity of the indices in predicting English literacy outcomes as seen in Tables A1-2.

Estimations are also run including elements individually. The rationale for this is that the first component in a principal components analysis describes less than 30% of the variation in the

[20] However, an African language ORF pre-test was not administered for the grade 6 sample. In estimating grade 6 African language ORF scores, a grade 6 English ORF score for each student is used as a pre-test control.

underlying factor, and inter-item correlations across elements are low. Individual elements may be predictive of outcomes, but this would not be observed when aggregating across elements. A striking result is how few of the individual 114 rubric elements are statistically significantly positively related to selected literacy outcomes. This is evident in Table 7, which summarises the percentage of elements under each dimension that are positive and significant (P*), positive but not significant (P), negative and significant (N*) and negative but not significant (N). Only 13% (15 of 114 elements) of the rubric elements are positively and significantly associated with grade 6 written literacy test outcomes using a full set of controls – only 12% if we include pre-test scores in a value-added model. Elements related to the effective development and deployment of human resources are most likely to yield positive and significant associations in value-added estimations of grade 6 literacy. For value-added estimates of grade 3 ORF outcomes, positive and significant associations are more commonly identified for elements defined under the 'use of time index' and presence of knowledge resources. A striking result however, which confirms a point made in earlier school effectiveness studies in South Africa (Gustafsson, 2007) is how often elements are not significant or even negatively correlated. Up to 10% of all 114 elements are statistically significantly negatively associated with literacy outcomes. While this may be surprising, the identification of negative associations is not uncommon in the international literature on 'leadership effects' (Robinson et al., 2008, p. 655).

Multivariate estimation of intermediate outcomes

Estimates of four intermediate outcomes yield stronger and more consistent associations compared with estimates of student literacy outcomes, as seen in Table 8. The human resources index as well as the strategic resources index are positively associated with coverage of work in the best grade 6 learners' exercise or workbooks (expressed in centiles), and the percentage of educators in the school indicating that their immediate manager (i.e. head of department or HoD) checks the amount of work they have covered in the curriculum at least twice a week[21]. A one standard deviation increase in the strategic resources index, for example, is associated with an 8 percentage point increase in the percentage of teachers who report their curriculum coverage is checked, and a 15 percentage point jump in exercise or workbook coverage.

The larger coefficient of 0.2 standard deviations on the human resource and strategic resource indices in estimates of teacher engagement is noteworthy. This suggests that stronger management and leadership processes may matter for the extent to which teachers enjoy their work, feel connected to their jobs and the school organisation. However, these findings are treated with caution.[22] When it comes to predicting a substantive measure of school functionality (the percentage of teachers who are identified as present and teaching in the classroom during school observations) none of the indices are significant and positive. One would expect higher teacher engagement levels to be met with evidence of more teaching in the classroom.

[21] The question asked of educators was "How often does your HoD (i.e head of department) in this school check to see how much of the curriculum you have taught?"

[22] This is an area of research that is worth further investigation especially in determining whether this reflects a real association or is driven response bias patterns in self-reported responses to perceptions on leadership, management and own engagement with work.

Table 5: Estimating grade 6 literacy outcomes using levels (PCA approach - Positively loading items - to combine rubric elements)

	Estimating Grade 6 English literacy (reading comprehension and vocabulary test, z-scores)				*Estimating Grade 6 English Oral Reading Fluency (z-scores)*				*Estimating Grade 6 African language oral reading fluency (% Words Read Correctly Per Minute)*			
	Levels		**Value-added**		**Levels**		**Value-added**		**Levels**		**Value-added**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Material Resources Time (z-score)	0.05	0.06	0.01	0.00	0.00	-0.01	0.01	0.00	0.16	0.14	0.30	0.21
	(0.08)	(0.06)	(0.02)	(0.02)	(0.08)	(0.05)	(0.02)	(0.02)	(0.67)	(0.55)	(0.54)	(0.33)
Pre-test control*			1.160***	1.134***			0.924***	0.895***			0.404***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.80	0.00	0.20	0.77	0.77	0.00	0.16	0.48	0.59
Material Resources Text Presence (z-score)	0.181**	-0.06	0.02	0.01	0.0987*	-0.09	-0.01	-0.0549*	-0.13	0.07	-1.003*	0.35
	(0.07)	(0.06)	(0.02)	(0.02)	(0.06)	(0.07)	(0.02)	(0.03)	(0.67)	(0.75)	(0.56)	(0.48)
Pre-test control*			1.158***	1.134***			0.925***	0.893***			0.410***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.02	0.25	0.80	0.80	0.01	0.20	0.77	0.77	0.00	0.16	0.48	0.59
Material Resources Text Use (z-score)	0.05	-0.03	-0.03	-0.03	0.04	-0.01	-0.02	-0.02	0.12	0.44	-0.41	0.38
	(0.11)	(0.08)	(0.03)	(0.03)	(0.09)	(0.07)	(0.03)	(0.03)	(0.82)	(0.61)	(0.73)	(0.38)
Pre-test control*			1.162***	1.134***			0.925***	0.896***			0.405***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.80	0.00	0.19	0.77	0.77	0.00	0.17	0.48	0.59
Knowledge Resources (z-score)	0.06	0.00	0.02	0.02	0.03	0.01	0.00	0.01	-0.15	0.42	-0.38	0.46
	(0.10)	(0.06)	(0.02)	(0.02)	(0.08)	(0.06)	(0.03)	(0.02)	(0.66)	(0.65)	(0.63)	(0.35)
Pre-test control*			1.159***	1.134***			0.924***	0.895***			0.404***	0.427***
			(0.02)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.80	0.00	0.20	0.77	0.77	0.00	0.17	0.48	0.59
Human Resources (z-score)	0.01	0.181**	0.0783**	0.100***	-0.01	0.162**	0.01	0.05	1.03	1.066*	1.527**	0.727*
	(0.07)	(0.06)	(0.03)	(0.03)	(0.06)	(0.07)	(0.03)	(0.03)	(0.66)	(0.57)	(0.57)	(0.39)
Pre-test control*			1.163***	1.126***			0.924***	0.889***			0.409***	0.425***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.26	0.80	0.81	0.00	0.21	0.77	0.77	0.01	0.17	0.49	0.59
	-0.03	0.10	0.02	0.03	-0.10	0.07	0.01	0.04	0.80	1.083*	1.657**	0.74

Strategic Resources (z-score)	(0.09)	(0.08)	(0.03)	(0.04)	(0.07)	(0.06)	(0.03)	(0.04)	(0.79)	(0.64)	(0.66)	(0.48)
Pre-test control*			1.161***	1.132***			0.925***	0.894***			0.413***	0.425***
			(0.02)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.80	0.01	0.20	0.77	0.77	0.00	0.17	0.50	0.59
Student observations	2541	2541	2379	2379	599	599	599	599	589	589	589	589
Controls												
Individual and home		X		X		X		X		X		X
Province		X		X		X		X		X		X
School		X		X		X		X		X		X

Source: Leadership for literacy dataset, 2017 - 60 schools.

Notes: Standard errors are in parentheses and clustered at the school level. Significant at *10% level, **5% level, ***1% level.

Each cell represents a regression of each individual index and the outcome variable.

The pre-test control in estimating A) grade 6 literacy is the z-score of the students' pre-test results on the same test. The pre-test control in estimating C) grade 6 African language oral reading fluency is the %WCPM of the student on the pre-test English ORF - no pre-test available in African language. In estimations of C., controls are also added to reflect the African language in question to account for differences in language structures.

Individual student characteristics include a measure of wealth and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student gets help with homework from someone at home, lives with his/her mother, lives with his/her father, the student's parents' employment status, rural vs. urban location, and whether the student speaks English at home `sometimes' or 'always' compared with 'never'. School controls, include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English.

Table 6: Estimating grade 3 literacy outcomes (PCA of positively loading items used to combine rubric elements)

	Estimating Grade 3 English Oral Reading Fluency (% of WCPM)				*Estimating Grade 3 African Oral Reading Fluency (% of WCPM)*			
	Levels		**Value-added**		**Levels**		**Value-added**	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Material Resources Time (z-score)	2.45	1.840**	1.338**	1.330**	3.708**	3.037**	2.040**	2.018**
	(1.64)	(0.89)	(0.64)	(0.55)	(1.41)	(1.30)	(0.72)	(0.72)
Pre-test control*			1.084***	0.989***			1.040***	0.992***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.01	0.32	0.78	0.80	0.02	0.18	0.73	0.74
Material Resources Text Presence (z-score)	-3.063**	-0.51	-0.11	0.47	-1.45	-0.50	0.55	0.75
	(1.32)	(0.91)	(0.61)	(0.62)	(1.35)	(1.53)	(0.85)	(0.92)
Pre-test control*			1.087***	0.993***			1.048***	0.997***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.02	0.31	0.78	0.79	0.00	0.17	0.73	0.74
Material Resources Text Use (z-score)	-3.407**	-1.741**	-0.33	0.23	-1.84	-2.777**	0.19	0.14
	(1.46)	(0.79)	(0.60)	(0.54)	(1.60)	(1.37)	(0.78)	(0.83)
Pre-test control*			1.085***	0.993***			1.047***	0.997***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.03	0.32	0.78	0.79	0.00	0.17	0.73	0.74
Knowledge Resources (z-score)	-0.66	-1.516*	0.01	-0.21	-1.48	-2.625**	-0.09	-0.34
	(1.81)	(0.86)	(0.62)	(0.52)	(1.75)	(1.27)	(0.76)	(0.71)
Pre-test control*			1.088***	0.990***			1.046***	0.995***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.00	0.32	0.78	0.79	0.00	0.18	0.73	0.74
Human Resources(z-score)	2.913*	-0.01	0.28	-0.33	1.80	-0.11	0.17	-0.12
	(1.57)	(0.84)	(0.64)	(0.56)	(1.66)	(1.28)	(0.78)	(0.73)
Pre-test control*			1.086***	0.992***			1.046***	0.996***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.02	0.31	0.78	0.79	0.00	0.17	0.73	0.74
Strategic Resources (z-score)	3.124**	2.485**	-0.63	-0.75	4.312**	3.872**	-0.63	-0.81
	(1.55)	(1.05)	(0.49)	(0.52)	(1.51)	(1.66)	(0.63)	(0.73)
Pre-test control*			1.096***	0.999***			1.051***	1.002***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.02	0.32	0.78	0.79	0.02	0.18	0.73	0.74
Student observations	631	631	631	631	631	631	631	631
Controls								
Individual and home		X		X		X		X
Province		X		X		X		X
School		X		X		X		X

Source: Leadership for literacy dataset, 2017 - 60 schools. **Notes:** Standard errors are in parentheses and clustered at the school level. Significant at *10% level, **5% level, **1% level. Each cell represents a regression of each individual index and the outcome variable. Individual student characteristics include a measure of wealth and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student lives with his/her mother, lives with his/her father, rural vs. urban location, whether the child has own story books and whether the student speaks the language of the test at home. School controls, include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English. In estimations of African language, controls are also added to reflect the African language in question to account for differences in language structures.

Table 7: The percentage of individual rubric elements within each leadership for literacy dimensions positively (P) or negatively (N) associated with literacy outcomes

	Estimating Grade 6 English literacy (reading comprehension and vocabulary test, z-scores)					*Estimating Grade 3 African Oral Reading Fluency (% of WCPM)*					*Estimating Grade 3 English Oral Reading Fluency (% of WCPM)*				
	Levels					*Levels*					*Levels*				
	P *	**P**	**N ***	**N**	**Total**	**P ***	**P**	**N ***	**N**	**Total**	**P ***	**P**	**N ***	**N**	**Total**
Material resources: Time (19 indicators)	11%	47%	5%	37%	100%	11%	53%	11%	26%	**100%**	11%	53%	16%	21%	**100%**
Material resources: Text (19 indicators)	21%	37%	21%	21%	100%	5%	32%	11%	53%	**100%**	0%	37%	5%	58%	**100%**
Material resources: Text use (9 indicators)	11%	33%	0%	56%	100%	22%	22%	22%	33%	**100%**	22%	22%	22%	33%	**100%**
Knowledge resource (18 indicators)	0%	50%	11%	39%	100%	11%	39%	17%	33%	**100%**	17%	28%	28%	28%	**100%**
Human resources (25 indicators)	20%	44%	8%	28%	100%	24%	40%	4%	32%	**100%**	24%	32%	4%	40%	**100%**
Strategic resources (24 indicators)	13%	79%	0%	8%	100%	25%	54%	0%	21%	**100%**	25%	54%	0%	21%	**100%**
All 114 indicators	13%	51%	8%	28%	100%	17%	42%	9%	32%	**100%**	17%	39%	11%	33%	**100%**
	Value-added					*Value-added*					*Value-added*				
	P *	**P**	**N ***	**N**	**Total**	**P ***	**P**	**N ***	**N**	**Total**	**P ***	**P**	**N ***	**N**	**Total**
Material resources: Time (19 indicators)	0%	53%	5%	42%	**100%**	16%	58%	0%	26%	**100%**	16%	42%	16%	26%	**100%**
Material resources: Text (19 indicators)	11%	32%	0%	58%	**100%**	11%	47%	16%	26%	**100%**	16%	63%	0%	21%	**100%**
Material resources: Text use (9 indicators)	22%	33%	0%	44%	**100%**	0%	56%	0%	44%	**100%**	0%	56%	0%	44%	**100%**
Knowledge resources (18 indicators)	6%	44%	11%	39%	**100%**	17%	39%	11%	33%	**100%**	17%	33%	11%	39%	**100%**
Human resources (25 indicators)	24%	52%	0%	24%	**100%**	4%	40%	8%	48%	**100%**	0%	36%	8%	56%	**100%**
Strategic resources (24 indicators)	13%	58%	0%	29%	**100%**	0%	42%	13%	46%	**100%**	0%	33%	17%	50%	**100%**
All 114 indicators	12%	47%	3%	38%	**100%**	8%	46%	9%	38%	**100%**	8%	42%	10%	40%	**100%**

Notes: P * = positive sign, and significant at 10% level. P = positive but not significant at 10% level. N* = negative sign, and significant at 10% level. N = negative sign, but not significant at 10% level. Each of the individual rubric elements are entered individually in each of the regressions. Coefficients on each rubric element are estimated using full set of controls (individual student, home background and school factors). Individual student characteristics include a measure of wealth and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student gets help with homework from someone at home, lives with his/her mother, lives with his/her father, the student's parents' employment status, rural vs. urban location, and whether the student speaks English at home `sometimes' or 'always' compared with 'never'. School controls, include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English.

Table 8: Relationship between LL indices and intermediate outcomes

		Coverage of work in exercise / workbooks (in centiles)			% of teachers present in classroom			% of educators indicating their curriculum coverage is monitored.			Index of teacher engagement (z-score)		
		Coef.	P-value	R2	Coef.	P-value	R2	Coef.	P-value	R2	Coef.	P-value	R2
Material Resources Time (z-score)	PCA	2.90	0.47	-0.03	1.47	0.52	0.02	2.56	0.17	0.51	0.06	0.14	0.24
Material Resources Text Presence (z-score)	PCA	5.18	0.35	-0.02	-0.50	0.91	0.02	-0.08	0.98	0.50	0.01	0.81	0.21
Material Resources Text Use (z-score)	PCA	-10.91	0.00	0.10	2.79	0.36	0.04	-0.69	0.81	0.50	-0.07	0.19	0.24
Knowledge Resources (z-score)	PCA	-2.52	0.57	-0.03	2.17	0.44	0.03	-1.21	0.62	0.50	0.02	0.62	0.21
Human Resources (z-score)	PCA	7.04	0.08	0.01	-0.09	0.97	0.02	5.28	0.10	0.54	0.20	0.00	0.45
Strategic Resources (z-score)	PCA	15.41	0.00	0.18	-3.36	0.12	0.04	7.60	0.01	0.58	0.21	0.00	0.43
SLM all combined (z-score)	Mean	8.52	0.03	0.04	0.25	0.93	0.02	2.84	0.21	0.51	0.16	0.00	0.37
Controls													
Student characteristics			X			X			X			X	
School characteristics			X			X			X			X	
Province			X			X			X			X	

Notes: N = 60 for all regressions. Each cell represents a regression of each individual index and the outcome variable.
Average student characteristics of grade 6 class include % overage, % who attended grade R, % who always or almost always speak English at home, % whose parents are not employed, % with own story books at home. School controls include average school wealth, class size of grade 6 class, English LOLT and low-fee paying. Cells are highlighted where p-values are less than 0.1.

V. Discussion

This paper described a process to attempt to construct the first available composite measure of leadership and management practices in South African schools that may be linked to learning, specifically the literacy outcomes of poorer students. Our rubric development approach centred on an articulated theoretical framework to construct measures that would capture a fuller SLM concept and in turn provide more predictive power of SLM than typically found in previous South African school effectiveness studies. While this research development is a positive move away from merely using proxies or indicator variables of SLM in education production function studies of learning, the results are disappointing. Overall, we find few linkages between six 'leadership for literacy' dimensions and grade 3 or 6 literacy outcomes. Only 12% of 114 individual rubric elements were positively and significantly associated with grade 6 literacy outcomes. Most of the time the expected relationships between evaluated areas of SLM competence or practice and outcomes are not found. This echoes earlier reflections in school effectiveness research in South Africa (Gustafsson, 2007; Van Staden and Howie, 2014).

On the one hand, this study raises some questions about whether leadership and management accounts for the lion's share of unexplained variation in learner performance across poorer schools in South Africa (see Crouch and Mabogoane, 1998). Even when more effort is given to quantifying a conceptual framework, the evidence discussed in this paper and the findings of in-depth case studies, imply a randomness to how better 'leadership for literacy' practices appear to be distributed across the school sample.

On the other hand, the results highlight the limitations and significant challenges to quantifying SLM constructs. Despite aims to strengthen the reliability of measures through adding triangulating responses and observational data, validity concerns taint the quantitative data collection process. Face validity was questioned when comparing results across in-depth case study findings and measures codified from data collected in quantitative instruments. One reason for this is that data collection instruments relied on close-ended questions to limit the cognitive load on fieldworkers and high-level judgements required. But this approach reduces the opportunity to probe through socially desirable responses. Triangulation was embedded into instrument design to address this limitation but may not be enough to compensate for using general fieldworkers rather than high-level researchers in gathering data on leadership and management practices in schools (Taylor et al., 2018). This is highlighted by the fact that predictive validity of the leadership for literacy indices did not improve notably when limiting index elements to those derived from observational or evidence-based data.

Despite these measurement concerns, our indexes were not completely lacking predictive validity. A 'human resource' dimension capturing the extent to which leaders develop, recognise and deploy expertise to building a literacy learning environment was most linked to grade 6 literacy outcomes, compared to 5 other dimensions. In estimating grade 3 literacy outcomes, the strongest linkages emerge with measures of the efficient use and allocation of time in the school for reading and language development. But the magnitude of these associations is still small. Stronger linkages are found with intermediate outcomes. Schools scoring higher on measures of effective deployment and utilisation of human or strategic resources have more evidence of curriculum management, curriculum coverage and teacher engagement with their work.

In furthering our work to quantify SLM in the South African context, what could be done? We believe we exhausted available opportunities to add more variation into the sample, using a purposeful sampling approach to identify best performing primary schools in three provinces and matching them

to underperforming pairs. However, there may be more variation in SLM practice *within* schools that could be exploited for such analyses. In a related paper we show that measures of teachers' perceptions and experiences of their managers and school leaders vary more within than across the 60 schools. Exploiting within school variation in SLM practice requires tests that can be compared across grades and subjects within a school. It is also possible that stronger linkages may be found with mathematics rather than literacy, as suggested by other studies (McCullough et al., 2016; Tavares, 2015).

VI. References

Arcia, G., Demas, A., 2013. What matters most for school finance: A framework paper (No. 2), Systems Approach for Better Education Results (SABER working paper series). Washington, D.C.

Arcia, G., Macdonald, K., Patrinos, H.A., Porta, E., 2015. School Autonomy and Accountability (No. 9), System Assessment and Benchmarking for Education Results, Systems Approach for Better Education Results (SABER working paper series). Washington, D.C. doi:10.1787/5k9h362kcx9w-en

Bloom, N., Lemos, R., Sadun, R., Scur, D., Van Reenen, J., 2014. The New Empirical Economics of Management. J. Eur. Econ. Assoc. 12, 835–876. doi:10.1111/jeea.12094

Bloom, N., Lemos, R., Sadun, R., Van Reenen, J., 2015. Does Management Matter in schools? Econ. J. 125, 647–674. doi:10.1111/ecoj.12267

Bloom, N., Van Reenen, J., 2010. New approaches to surveying organizations. Am. Econ. Rev. 100, 105–109. doi:10.1257/aer.100.2.105

Branch, G.F., Hanushek, E.A., Rivkin, S.G., 2013. School leaders matter. Educ. Next 63–69.

Branch, G.F., Hanushek, E.A., Rivkin, S.G., 2012. Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals (No. 17803), National Bureau of Economic Research Working Paper Series. Cambridge, MA. doi:10.3386/w17803

Bush, T., Glover, D., 2016. School leadership and management in South Africa: Findings from a systematic literature review. Int. J. Educ. Manag. 30, 211–213.

Camburn, E.M., Spillane, J.P., Sebastian, J., 2010. Assessing the Utility of a Daily Log for Measuring Principal Leadership Practice. Educ. Adm. Q. 46, 707–737. doi:10.1177/0013161X10377345

Carnoy, M., Ngware, M., Oketch, M., 2015. The Role of Classroom Resources and National Educational Context in Student Learning Gains: Comparing Botswana, Kenya, and South Africa. Comp. Educ. Rev. 59, 199–233. doi:10.1086/680173

Coelli, M., Green, D.A., 2012. Leadership effects: School principals and student outcomes. Econ. Educ. Rev. 31, 92–109. doi:10.1016/j.econedurev.2011.09.001

Condon, C., Clifford, M., 2012. How Rigorous Are Commonly Used Principal Performance Assessment Instruments? Qual. Sch. Leadersh. Issue Br.

Crawfurd, L., 2017. School Management and Public–Private Partnerships in Uganda. J. Afr. Econ. 26, 539–560. doi:10.1093/jae/ejx021

Crouch, L., Mabogoane, T., 2001. No Magic Bullets, just Tracer Bullets: The role of learning resources,

social advantage and education management in improving the performance of South African schools. Soc. Dyn. 27, 60–78. doi:10.1080/02533950108458704

Crouch, L., Mabogoane, T., 1998. When the residuals matter more than the coefficients: An educational perspective. Stud. Econ. Econom. 22, 1–14.

Department of Basic Education (DBE), 2016. Personnel Administrative Measures (PAM) Government Gazette No. 39684, 12 February 2016. Pretoria.

Fryer, R.G., 2017. Management and Student Achievement: Evidence from a Randomized Field Experiment (No. 23437), NBER Working Paper Series. Cambridge, MA.

Goldring, E., Cravens, X.C., Murphy, J., Porter, A.C., Elliott, S.N., Carson, B., 2009. The evaluation of principals: What and how do states and urban districts assess leadership? Elem. Sch. J. 110, 19–39.

Grissom, J.A., Kalogrides, D., Loeb, S., 2015a. Using Student Test Scores to Measure Principal Performance. Educ. Eval. Policy Anal. 37, 3–28.

Grissom, J.A., Loeb, S., Mitani, H., 2015b. Principal time management skills: Explaining patterns in principals' time use, job stress and perceived effectiveness. J. Educ. Adm. 53, 773–793. doi:ps://doi.org/10.1108/JEA-09-2014-0117

Gustafsson, M., 2007. Using the Hierarchical Linear Model To Understand School Production in South Africa. South African J. Econ. 75, 84–98. doi:10.1111/j.1813-6982.2007.00107.x

Gustafsson, M., 2005. The relationships between schooling inputs and outputs in South Africa: Methodologies and policy recommendations based on the 2000 SACMEQ dataset, SAQMEQ paper series. Botswana.

Hallinger, P., 2017. Surfacing a hidden literature. Educ. Manag. Adm. Leadersh. 46, 174114321769489. doi:10.1177/1741143217694895

Hallinger, P., 2008. Methodologies for studying school leadership: A review of 25 years of research using the principal instructioal management rating scale. Pap. Present. Annu. Meet. Am. Educ. Res. Assoc. New York. 48.

Harter, J.K., Schmidt, F.L., Hayes, T.L., 2002. Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. J. Appl. Psychol. doi:10.1037/0021-9010.87.2.268

Heck, R.H., Hallinger, P., 2005. The Study of Educational Leadership and Management: Where Does the Field Stand Today? Educ. Manag. Adm. Leadersh. 33, 229–244. doi:10.1177/1741143205051055

Hoadley, U., 2018. Leading for literacy: A review of the research, Research on Socio-Economic Policy Working Papers. Stellenbosch.

Hoadley, U., Christie, P., Ward, C., 2009. Managing to learn: Instructional leadership in South African secondary schools. Sch. Leadersh. 29, 373–389. doi:10.1080/13632430903152054

Hoadley, U., Ward, C., 2009. Managing to Learn: Instructional Leadership in South African Secondary Schools. HSRC press, Pretoria.

Horng, E.L., Klasik, D., Loeb, S., 2009. Principal Time-Use and School Effectiveness (School Leadership and Research Report No. 09-3). Standford, CA.

Howie, S., Combrink, C., Roux, K., Tshele, M., Mokoena, G.M., McLeod Palane, N., 2017. PIRLS Literacy 2016: South African Highlights Report. Pretoria.

Kahn, W.A., 1990. Psychological Conditions of Personal Engagement and Disengagement at Work. Acad. Manag. J. doi:10.5465/256287

Kotze, J., 2017. Social Gradients, Early Childhood Education and Schools Performing Above the Demographic Expectation: Empirical Insights into Educational Issues. Stellenbosch University.

Lee, M., Hallinger, P., 2012. National contexts influencing principals' time use and allocation: economic development, societal culture, and educational system. Sch. Eff. Sch. Improv. 23, 461–482.

Leithwood, K., Louis, K.S., Anderson, S., Wahlstrom, K., 2004. Review of research. How leadership influences student learning, Learning from Leadership Project. New York. doi:10.1007/978-90-481-2660-6

Lemos, R., Scur, D., 2017. Developing Management: An expanded evaluation tool for developing countries (No. 16/007), RISE working paper series.

McCullough, M., Lipscomb, S., Chiang, H., Gill, B., Cheban, I., 2016. Measuring school leaders' effectiveness: Final report from a multiyear pilot of Pennsylvania's framework for leadership, (REL 2016–106). Washington, D.C.

Peters, G.Y., 2014. The alpha and the omega of scale reliability and validity. Eur. Heal. Psychol. 1, 54–72. doi:10.31234/osf.io/h47fv

Piedmont, R.., 2014. Inter-item Correlations., in: Michalos, A.C. (Ed.), Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.

Robinson, V.M.J., Lloyd, C.A., Rowe, K.J., 2008. The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. Educ. Adm. Q. 44, 635–674. doi:10.1177/0013161X08321509

Scur, D., 2017. Organization matters: The causes and consequences of organizational choies in the private and public sectors. University of Oxford.

Shepherd, D., 2016. The open door of learning - access restricted: School effectiveness and efficiency across the South African education system. Stellenbosch University, Stellenbosch.

Shepherd, D.L., 2011. Constraints to school effectiveness: What prevents poor schools from delivering results? (No. 05/11), Stellenbosch Economic Working Papers. Stellenbosch.

Spaull, N., 2013a. South Africa's Education Crisis: The quality of education in South Africa 1994-2011. (Report commissioned by the CDE). Johannesburg. doi:10.1007/BF03217477

Spaull, N., 2013b. Poverty & Privilege: Primary School Inequality in South Africa. Int. J. Educ. Dev. 33, 436–447.

Stein, M.K., Nelson, B.S., 2003. Leadership content knowledge. Educ. Eval. Policy Anal. 25, 423–228.

Tavares, P.A., 2015. The impact of school management practices on educational performance: Evidence from public schools in São Paulo. Econ. Educ. Rev. 48, 1–15. doi:10.1016/j.econedurev.2015.05.002

Taylor, N., Gamble, J., Spies, M., Garisch, C., 2013. Chapter 4: School leadership and management, in: Taylor, N., van der Berg, S., Mabogoane, T. (Eds.), Creating Effective Schools: Report of South

Africa's National School Effectiveness Study. Pearson Education, Cape Town.

Taylor, N., Prinsloo, C., 2005. The Quality Learning Project Lessons for High School Improvement in South Africa. Johannesburg.

Taylor, N., Wills, G., Hoadley, U., 2018. Addressing the "leadership conundrum" through a mixed methods study of school leadership for literacy. Stellenbosch.

Taylor, S., 2011. Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study (No. 10/11), Stellenbosch Economic Working Papers. Stellenbosch.

Van der Berg, S., 2008. How effective are poor schools? Poverty and educational outcomes in South Africa. Stud. Educ. Eval. 34, 145–154. doi:10.1016/j.stueduc.2008.07.005

Van der Berg, S., Louw, M., 2006. Unravelling the mystery: Understanding South African schooling outcomes in regional context, in: Conference for the Study of African Economics, Oxford University. pp. 1–32.

Van Staden, S., Howie, S., 2014. Reflections on Creemers' Comprehensive Model of Educational Effectiveness for Reading Literacy: South African Evidence from PIRLS 2006. Perspect. Educ. 32, 172–192.

Wills, G., 2017. What do you mean by 'good'? The search for exceptional primary schools in South Africa' s no-fee school system (No. WP16/2017), Stellenbosch Economics Working Paper Series.

Witziers, B., Bosker, R.J., Krüger, M.L., 2003. Educational leadership and student achievement: The elusive search for an association. Educ. Adm. Q. 39, 398–425. doi:10.1177/0013161X03253411

Table A 1: Estimating grade 6 literacy outcomes using levels (PCA approach – evidence-based items)

	Estimating Grade 6 English literacy (reading comprehension and vocabulary test, z-scores)				*Estimating Grade 6 English Oral Reading Fluency (z-scores)*				*Estimating Grade 6 African language oral reading fluency (% Words Read Correctly Per Minute)*			
	Levels		**Value-added**		**Levels**		**Value-added**		**Levels**		**Value-added**	
	Model 1	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**
Material Resources Time (z-score)	0.141*	0.08	0.01	0.01	0.05	0.03	0.02	0.02	-0.05	0.26	-0.09	0.11
	(0.08)	(0.06)	(0.02)	(0.02)	(0.07)	(0.06)	(0.03)	(0.03)	(0.68)	(0.56)	(0.67)	(0.40)
Pre-test control*			1.159***	1.133***			0.923***	0.895***			0.404***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.01	0.25	0.80	0.80	0.00	0.20	0.77	0.77	0.00	0.16	0.48	0.59
Material Resources Text Presence (z-score)	0.141*	-0.06	0.02	0.03	0.0613	-0.10	-0.02	-0.0625*	-0.24	0.07	-0.87	0.38
	(0.07)	(0.07)	(0.03)	(0.03)	(0.07)	(0.07)	(0.03)	(0.04)	(0.74)	(0.76)	(0.57)	(0.52)
Pre-test control*			1.158***	1.135***			0.926***	0.893***			0.407***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.01	0.25	0.80	0.80	0.00	0.20	0.77	0.77	0.00	0.16	0.48	0.59
Material Resources Text Use (z-score)	0.06	-0.03	-0.04	-0.03	0.05	-0.01	-0.02	-0.02	0.13	0.40	-0.44	0.36
	(0.11)	(0.08)	(0.03)	(0.03)	(0.09)	(0.07)	(0.03)	(0.03)	(0.83)	(0.61)	(0.73)	(0.39)
Pre-test control*			1.163***	1.134***			0.925***	0.895***			0.405***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.80	0.00	0.20	0.77	0.77	0.00	0.17	0.48	0.59
Knowledge Resources (z-score)	0.11	0.01	0.03	0.0340*	0.03	0.01	0.00	0.01	0.04	0.48	-0.50	0.43
	(0.11)	(0.06)	(0.02)	(0.02)	(0.08)	(0.06)	(0.03)	(0.02)	(0.66)	(0.63)	(0.66)	(0.37)
Pre-test control*			1.158***	1.135***			0.923***	0.896***			0.405***	0.427***
			(0.02)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.01	0.25	0.80	0.80	0.00	0.19	0.77	0.77	0.00	0.17	0.48	0.59
Human Resources(z-score)	-0.08	0.117**	0.0561**	0.0764**	-0.05	0.119	0.01	0.04	0.37	0.73	1.01	-0.11
	(0.08)	(0.05)	(0.02)	(0.03)	(0.07)	(0.07)	(0.03)	(0.04)	(0.62)	(0.81)	(0.64)	(0.66)
Pre-test control*			1.166***	1.131***			0.925***	0.892***			0.408***	0.427***
			(0.03)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.00	0.25	0.80	0.81	0.00	0.20	0.77	0.77	0.00	0.17	0.48	0.59

Strategic Resources (z-score)	0.300***	0.0974*	0.01	0.00	0.218***	0.09	0.05	0.05	0.08	0.18	-1.279*	-0.25
	(0.08)	(0.06)	(0.02)	(0.02)	(0.06)	(0.06)	(0.03)	(0.04)	(0.68)	(0.68)	(0.66)	(0.54)
Pre-test control*			1.159***	1.133***			0.915***	0.893***			0.415***	0.427***
			(0.02)	(0.02)			(0.04)	(0.04)			(0.03)	(0.03)
Adjusted R-squared	0.05	0.25	0.80	0.80	0.04	0.20	0.77	0.77	0.00	0.16	0.49	0.59
Observations	2541	2541	2379	2379	599	599	599	599	589	589	589	589
Controls												
Individual and home		X		X		X		X		X		X
Province		X		X		X		X		X		X
School		X		X		X		X		X		X

Source: Leadership for literacy dataset, 2017 - 60 schools.

Notes: Standard errors are in parentheses and clustered at the school level. Significant at *10% level, **5% level, ***1% level.

Each cell represents a regression of each individual index and the outcome variable.

The pre-test control in estimating A) grade 6 literacy is the z-score of the students' pre-test results on the same test. The pre-test control in estimating C) grade 6 African language oral reading fluency is the %WCPM of the student on the pre-test English ORF - no pre-test available in African language. In estimations of C., controls are also added to reflect the African language in question to account for differences in language structures.

Individual student characteristics include a measure of wealth and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student gets help with homework from someone at home, lives with his/her mother, lives with his/her father, the student's parents' employment status, rural vs. urban location, and whether the student speaks English at home `sometimes' or 'always' compared with 'never'. School controls, include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English.

Table A 2: Estimating grade 3 literacy outcomes (PCA of evidence-based elements)

	Estimating Grade 3 English Oral Reading Fluency (% of WCPM)				*Estimating Grade 3 African Oral Reading Fluency (% of WCPM)*			
	Levels		**Value-added**		**Levels**		**Value-added**	
	Model 1	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**	**Model 1**	**Model 3**
Material Resources Time (z-score)	1.956	1.014	1.001*	0.915	1.148	1.363	0.984	1.289
	(1.24)	(0.90)	(0.52)	(0.61)	(1.31)	(1.27)	(0.75)	(0.82)
Pre-test control*			1.085***	0.991***			1.046***	0.996***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.01	0.32	0.78	0.80	0.00	0.17	0.73	0.74
Material Resources Text Presence (z-score)	-2.214	-0.496	0.0693	0.413	-1.01	-0.127	0.625	0.797
	(1.37)	(1.05)	(0.58)	(0.59)	(1.35)	(1.77)	(0.83)	(0.92)
Pre-test control*			1.089***	0.992***			1.048***	0.997***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.01	0.31	0.78	0.79	0.00	0.17	0.73	0.74
Material Resources Text Use (z-score)	-3.396**	-1.769**	-0.311	0.219	-1.846	-2.760**	0.181	0.108
	(1.44)	(0.77)	(0.59)	(0.54)	(1.55)	(1.34)	(0.78)	(0.83)
Pre-test control*			1.085***	0.993***			1.047***	0.997***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.03	0.32	0.78	0.79	0.00	0.17	0.73	0.74
Knowledge Resources (z-score)	-1.405	-1.307	-0.283	-0.332	-1.893	-2.260*	-0.347	-0.503
	(1.80)	(0.83)	(0.62)	(0.51)	(1.78)	(1.25)	(0.80)	(0.68)
Pre-test control*			1.087***	0.990***			1.045***	0.994***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.00	0.32	0.78	0.79	0.00	0.17	0.73	0.74
Human Resources (z-score)	4.451**	0.061	0.848	-0.076	2.191	0.0418	0.417	0.221
	(1.39)	(0.78)	(0.65)	(0.61)	(1.55)	(1.26)	(0.78)	(0.78)
Pre-test control*			1.078***	0.991***			1.045***	0.996***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.05	0.31	0.78	0.79	0.01	0.17	0.73	0.74
Strategic Resources (z-score)	-1.228	0.891	-0.785	-0.423	-0.744	1.654	-0.835	-0.455
	(1.59)	(0.90)	(0.60)	(0.47)	(1.51)	(1.49)	(0.70)	(0.72)
Pre-test control*			1.087***	0.993***			1.046***	0.998***
			(0.03)	(0.03)			(0.03)	(0.04)
Adjusted R-squared	0.00	0.32	0.78	0.79	0.00	0.17	0.73	0.74
Observations	631	631	631	631	631	631	631	631
Controls								
Individual and home		X		X		X		X
Province		X		X		X		X
School		X		X		X		X

Source: Leadership for literacy dataset, 2017 - 60 schools. **Notes:** Standard errors are in parentheses and clustered at the school level. Significant at *10% level, **5% level, **1% level. Each cell represents a regression of each individual index and the outcome variable. Individual student characteristics include a measure of wealth and its square; the student's age, gender and whether the student attended grade R. Home controls include an indicator variable for whether the student lives with his/her mother, lives with his/her father, rural vs. urban location, whether the child has own story books and whether the student speaks the language of the test at home. School controls, include average school wealth, class size, low-fee vs. no-fee status and indicators for whether a schools' language of learning or teaching in the foundation phase is English. In estimations of African language, controls are also added to reflect the African language in question to account for differences in language structures.

Table A 3: Explained variation of the first principal component and inter-item correlations for 6 Leadership for literacy indices

Index combination approach		Material resources: Time	Material resources: Presence of text	Material resources: Use of text	Human Resources	Symbolic resources	Strategic resources
1. PCA all elements	Inter-item corr	0.05	0.19	0.11	0.08	0.06	0.13
	N items	19	19	9	25	20	24
	PCA - rho	0.12	0.33	0.29	0.17	0.18	0.21
2. PCA - +ve loading elements	Inter-item corr	0.07	0.20	0.14	0.13	0.07	0.15
	N items	14	18	8	16	18	22
	PCA - rho	0.16	0.33	0.31	0.27	0.19	0.23
3. PCA 'evidence-based' elements	Inter-item corr	0.08	**0.41**	0.12	0.16	0.09	0.14
	N items	8	**8**	9	3	8	8
	PCA – rho	0.24	**0.54**	0.29	0.46	0.34	0.26

Notes: Ideally, the. average inter-item correlation for a set of items should be between 0.20 and 0.40 (Piedmont, 2014). PCA-rho reflects that percentage of variation explain on the first component in a principal components analysis.