December

2018

RESEP WORKING PAPER

Addressing the 'leadership conundrum' through a mixed methods study of school leadership for literacy

Nick Taylor, Gabrielle Wills and Ursula Hoadley

RESEP

Research on Socio-Economic Policy

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Department of
ECONOMICS

Addressing the 'leadership conundrum' through a mixed methods study of school leadership for literacy

Nick Taylor, Gabrielle Wills and Ursula Hoadley[1]

30 November 2018

Abstract

This paper explores methodological insights from a mixed methods study that aims to understand how school leaders promote literacy development in their schools. The study findings consider both the complementarities and challenges of the qualitative and quantitative approaches to measuring leadership practices and their linkages with learning across schools. We begin by identifying a conundrum in school leadership and management (SLM) research – strong effects found in qualitative studies and weaker effects in quantitative studies. From the literature we identify some of the central challenges that account for these differences. We then show how these challenges were and were not addressed in the mixed method research we conducted in an SLM study of South African primary schools in challenging contexts. We consider why the central aim of the study – to develop a scalable instrument for measuring SLM – remains elusive.

Keywords: Mixed methods, school leadership and management, student achievement, South Africa, challenging contexts

[1] Nick Taylor (ntaylor@jet.org.za) is based at the Joint Educational Trust (JET); Gabrielle Wills (gabriellewills@sun.ac.za) is a researcher with Research on Socio-Economic Policy (ReSEP) and Ursula Hoadley (uk.hoadley@uct.ac.za) is a professor at the University of Cape Town's (UCT) education faculty.

1. Introduction

The broader study in which this paper is located aimed to understand how school leaders promote literacy development in their schools. This aim was to be realised through an in-depth investigation of the realities and possibilities of the role of school leadership and management (SLM) practices in improving reading instruction under circumstances which frame schooling for South African children from poor homes. Students in 'no-fee'[2] schools constitute some 70% of their age cohort, and the overwhelming majority failed to attain the Low International Benchmark in the 2016 iteration of the Progress in Reading Literacy Study tests (Mullis et al., 2017).

Four objectives motivated the broader study:

i.   Identify the number of exceptional rural and township primary schools in South Africa.
ii.   Gain new insights into school leadership and management (SLM) practices in high achieving schools relative to average or low-achieving schools in challenging contexts using case studies.
iii. Develop a scalable SLM instrument that captures the practices and behaviours of school leaders and managers in challenging contexts in South Africa.
iv.   Establish predictive validity - how predictive is this SLM instrument of academic achievement in these schools?

The present paper is mainly methodological in nature, grappling with methodological challenges in meeting objectives ii – iv. While reflecting important insights into some of the leadership practices observed in the case study schools, it is primarily concerned with the relative strengths and weaknesses of quantitative and qualitative methods, respectively, in informing the research questions and the ways in which they complement each other. It also attempts to provide explanations for areas where the two approaches appear to contradict each other. Since the main purpose of the paper is methodological, we do not provide extensive contextual details on the South African school system or the powers and functions of school leaders, except where background information is required to understand a particular substantive finding.

2. The leadership conundrum

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. Robinson et al (2008), commenting on international research by Edmonds (1979) and Maden (2001), reflects that in most case studies of school turnaround, rejuvenation is attributed to changes in leadership. New principals are

---

[2] South African schools are classified into quintiles according to the level of poverty of their feeder communities and school infrastructure. Parents pay no fees in the poorest 60% of schools (quintiles 1-3), which formed the object of the present study.

responsible for reviving dysfunctional schools to the point that academic achievements improve considerably to meet or even exceed learning benchmarks. In a review of case studies on school leadership and how it influences student learning, Leithwood et al. (2004: 7) identify that:

> *Indeed, there are virtually no documented instances of troubled schools being turned around without intervention by a powerful leader. Many other factors may contribute to such turnarounds, but leadership is the catalyst.*

In contrast to these positive findings, studies using quantitative data designs often contradict the heroic value placed on leadership and management and its ability to generate student achievement. Numerous reviews exist of quantitative studies of educational leadership effects on school outcomes, specifically student achievement. These reviews broadly divide studies into those that consider overall leadership effects and those that explore the effects of specific leadership practices. The overwhelming consensus from these studies is that in general leadership effects are weak and small (Hallinger and Heck, 1996; Leithwood et al., 2004; Robinson et al., 2008). For example, reporting on evidence from a meta-analysis of 37 international studies by Witziers, an average leadership effect on student outcomes in the form of a z-score was only 0.02 which reflects no or very weak impact (Witziers et al., 2003). There are more recent large quantitative studies which find educationally significant principal effects but the estimation of effect sizes varies notably depending on estimation model assumptions (Branch et al., 2012; Grissom et al., 2015).

Addressing the leadership conundrum

At least three methodological difficulties are implicated in producing the contradictory findings observed across qualitative and quantitative research in educational leadership and management and in explaining the comparatively smaller effects identified in quantitative studies.

First, different sampling strategies adopted in the qualitative and quantitative disciplines are identified as a key reason for the contradictory evidence (Robinson et al., 2008). Using random sampling techniques, large quantitative studies measure "average" leadership effects. However, by grouping together schools across a spectrum of needs that have divergent leadership skills, Leithwood and colleagues argue that such studies systematically underestimate leadership effects in schools where it is likely to be of greatest value (Leithwood et al., 2004). By contrast it is from these very schools, in greatest need of leadership, that qualitative studies deduce the importance of the role of SLM for school functioning. We attempted to address this issue by purposively sampling better performing schools and matching them with lower performing schools with similar demographic features. In this way we aimed to introduce maximal variation into the sample in terms student performance (and potentially leadership/management practice) that may exist in these schools.

A second possible explanation for small or insignificant effects found in quantitative studies regards the validity and reliability of instruments used to measure leadership and

management in quantitative surveys (Hallinger and Heck, 1996). Our response to this challenge was to commence instrument design with the development of a strong theoretical framework, followed by an instrument design process with various stages of item writing and piloting to foster content validity.

Finally, the effects of school leadership may not exhibit in teaching processes and learning outcomes because of what Pritchett et al. (2013) have likened as 'isomorphic mimicry', where leaders go through the motions of compliance with policy or known best practice, but whose actions fail to achieve the desired outcomes because of poor implementation. Such behaviour may be motivated by 'malicious compliance', characterised by leaders pretending to adopt policy, but not following through to practice, or it may be due to ignorance on the part of leaders to fully understand both the letter and spirit of the policy. Closely related to isomorphic mimicry is the production of 'socially acceptable' responses, where respondents tell the interviewer what they believe the latter wants to hear, or what they perceive to be accurate (Mertler, 2019).

In response to this challenge, investigating the extent to which isomorphic mimicry is present in the policies and practices of the case study schools was one of the explicit aims of the qualitative component, where techniques such as triangulation, and semi-structured, probing interviews were employed in an attempt to get beneath the surface of intentions and claims and understand the link between policy intentions and SLM practices.

3.   Research method

A mixed method approach

The study used a mixed method design. Part of the reason for this was alluded to in the discussion of the leadership conundrum above, but it was also motivated by familiarity with the difficulty of putting a finer point of the residual found in school effectiveness studies, often attributed to school leadership and management. We considered a mixture of quantitative and qualitative approaches appropriate to detailed explorations of leadership at the micro level that could then potentially be converted into quantifiable factors for survey use.

The field of mixed methods research has advanced considerably over the last few decades, and there are a range of established mixed method designs and typologies (Creswell, 2003; Onwuegbuzie and Johnson, 2004; Tashakkori and Teddlie, 1998, 2003) indicating the ways in which qualitative and quantitative approaches potentially combine. Although these models are useful in identifying designs and clarifying approaches, in practice such design options "are neither exclusive nor singular because actual mixed methods studies are often much more complex than any single-design alternative can adequately represent" (Jang et al., 2008: 224). Further, we found the application of a mixed methods strategy very challenging, with the need to adapt and combine models as we proceeded. This occurred not least in relation to the difficulty of predicting different time frames for the different types of research.

In terms of the aforementioned models, our study is best described as *sequential* (Tashakkori and Creswell, 2007) in the development of theory, and the design and development of instruments. The qualitative case studies notably fed into the development of the survey instruments. The study was *concurrent* (or parallel) in the collection and analysis of data, in that quantitative and qualitative strands functioned separately at these phases of the research. This allowed us to verify findings by utilizing both qualitative and quantitative strands. Further, results (and non-results) from the survey were clarified with contextually specific and detailed cases and an attempt was made to synthesize results from both strands to understand better our research problem and issues of measurement.

Our study can also be described as *integrated* (Caracelli and Greene, 1997) in that ''mixing'' occurred at different points: our research questions were aligned with both methods, preliminary analysis of each phase informed the data collection of subsequent phases and a 'quantification' of the qualitative data in the final analyses for purposes of comparison also indicates integration in the research approach.

Both quantitative and qualitative research perspectives have yielded important insights into the relationship between educational practices and performance but each of these lenses, on its own, leaves questions unanswered (Deaton, 2010; Deaton and Cartwright, 2018). In investigating school- or teacher-focussed interventions, strong experimental designs are best suited to establishing beyond reasonable doubt the effects of certain programmes on learning outcomes (Fleisch et al., 2016; Fleisch and Schöer, 2012; Piper, 2009). However, such studies often leave us wondering *how* these effects were achieved. Qualitative case studies, on the other hand, are better suited to understanding the generative mechanisms for changes in teaching and learning but beg the question as to whether the observed practices alone are likely to lead to similar changes in different schools, or whether the observed changes are the result of some idiosyncrasy in the case study schools.

Mixed method designs set out to extract optimal benefit from both research approaches. A South African example is afforded by the Early Grade Reading Study in which a mixed methods impact evaluation design is used to quantitatively test the effectiveness of two intervention models aimed at assisting teachers with more effective reading instruction, and qualitatively uncovering the mechanisms of change in each (Kotze et al., 2018; Taylor et al., 2017). The study described below is not an intervention but uses both quantitative and qualitative approaches in investigating the effects of different SLM practices on reading performance. However, the success of this study would depend on how well we could address methodological shortfalls outlined above that explain the leadership conundrum.

Theoretical frame

The study commenced with a review of the literature related to leadership and the teaching and learning of reading. The objective of the review was to draw out a set of factors relevant to leadership for literacy. Despite the suggestion by Crouch and Mabogoane (2001) of the importance of school management in explaining learning in the South African context, few local studies have successfully quantified key SLM factors implicated in improved teaching and learning. Some preliminary work has been done in this regard (Hoadley et al., 2009;

Hoadley and Galant, 2015; Taylor et al., 2013), but there is still limited understanding of which SLM practices contribute to or detract from school functionality, particularly with respect to producing learning outcomes in South Africa (Bush and Heystek, 2006).

In response, the study described in the present paper was dedicated to the measurement of school leadership and management and understanding its effects on student achievement. The literature review identified four kinds of resources available to school leaders in promoting literacy in the school: leaders' understanding of literacy and how it is learnt (knowledge resources), the recruitment and deployment of educators within the school (human resources), the material resources required for reading (material resources) and the extent to which these resources are mobilised in driving a coherent literacy programme (strategic resources) (Hoadley, 2018). These four sets of resources constituted an analytic framework for the study, and refined the research question guiding the study:

To what extent do school leaders develop and deploy resources (knowledge, human, material and strategic) to best advantage in promoting the teaching and learning of reading throughout the school?

Ten specific questions probing the extent to which these resources are present and utilised in the sample schools were then formulated (see Table 1).

Some of the practices to which these questions refer are more directly under the control of school leaders than others. For example, with reference to Q3, policies directed from the national or provincial departments of education may inhibit the discretion of school principals to recruit teachers with particular skill sets. With respect to Q4, on the other hand, the principal or other members of the leadership team may have more leeway in identifying teachers within the school with particular strengths and structure opportunities for them to assist their peers who may be lacking in these pedagogical skills. As we see below, the data in this study suggest that the most skilful leaders are those who bend restrictive external forces to serve the best interests of the school.

Sample

The quantitative approach to the project was embedded in a sampling process with strong qualitative nuances – the matched pairs design. The schools were purposively selected from challenging contexts, namely 'no-fee' schools in township and rural settings. Through the matched pairs design we also intentionally aimed to add as much variation into the sample in terms of student performance (and potentially leadership/management practice) that may exist in these schools. We engaged in a rigorous process to identify the best possible high-performing no-fee schools in three provinces using system-wide testing data in the form of the Annual National Assessments (ANA). (This process is described in detail in Wills, 2017).

Table 1: Research questions derived from the literature review

Resource	Indicators	Research questions
Knowledge resources	Extent to which school leaders understand the value and technology of teaching reading. The extent to which this understanding is shared by school leaders and with teachers across the school.	Q1: Are school leaders knowledgeable about teaching reading? Q2: Is this knowledge shared with educators across the school?
Human resources	The extent to which expertise in reading instruction is used to recruit and promote educators. The extent to which existing expertise is recognised and utilised across the school. The extent to which expertise is developed.	Q3: Is the recruitment, management, and deployment of staff in the school related to the promotion of reading instruction? Q4: To what extent is expertise in teaching reading recognised in individual teachers and used advantageously throughout the school? Q5: Are educators provided opportunities to collaborate and share expertise in the interests of improving reading instruction
Strategic resources	The extent to which there are programmes and practices in the school geared towards the improvement of reading instruction and outcomes.	Q6: Are there programmes and practices in the school that are geared towards the improvement of reading instruction and outcomes?
Material resources	The extent which time is used effectively for reading instruction. The extent to which textual resources are procured, deployed, and utilised for effective reading instruction.	Q7: Does the school prioritise the acquisition of high quality textual resources to support a programme of reading? Q8: Are the texts utilized optimally?

Due to the possible irregularities in ANA testing and marking processes, school performance on ANA was corroborated with a large dataset we collected of recommended 'good' schools from a host of sources (district officials, school principals and administrative clerks, education related non-governmental organisations, unions, other stakeholders, secondary schools performing well in the school-leaving examination called the National Senior Certificate). We also added into the sample 5 low-fee schools to add additional variation. A total sample of 30 better performing schools in poor communities were matched to 30 lower ANA performing schools located in similar geographic locations.

Selection of 8 case study schools was done after the pre-test survey of the full sample of 60 schools, including the administration of various literacy and reading tests. The first stage in

this process was to select the four best performing schools in terms of Grade 3 and 6 literacy scores on the tests described below. These high-performing case study schools were then matched with four schools performing worse in the literacy tests, but with sufficient overlap in the socio-economic status of the tested Grade 6 class. The success of this sampling strategy in ensuring performance variation across the spectrum of school studies is illustrated in Figure 1.
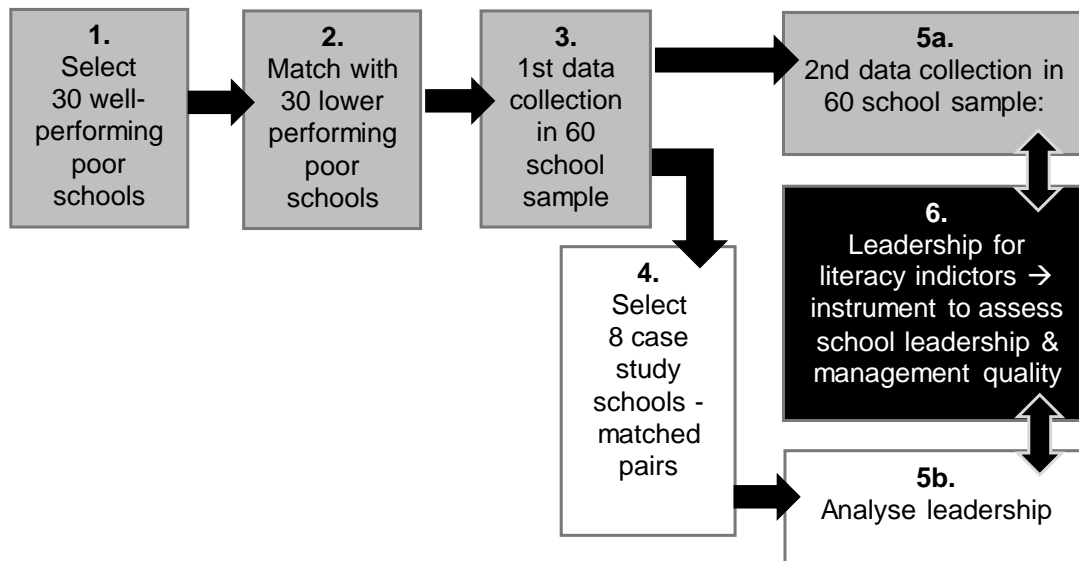
Figure 1: Grade 6 English literacy post-test class median score vs. pre-test class median score
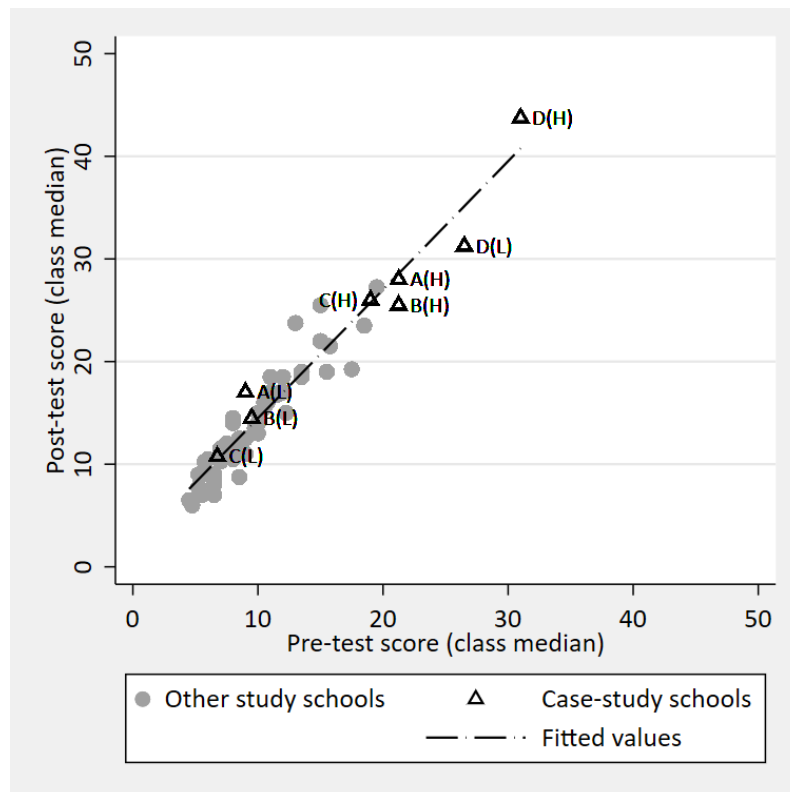
Design and method

To assess the extent of convergence or divergence across qualitative and quantitative approaches to measuring leadership for literacy competencies, a convergent parallel mixed methods design (Creswell, 2013) was used, which we discussed above as concurrent or sequential at certain stages and integrated at others. Qualitative and quantitative data were collected in parallel (sequentially), analysed separately (concurrently), and then merged (integrated). The research design is summarised in Figure 2.

Figure 2: Leadership for Literacy research design

1. Select 30 well-performing poor schools	2. Match with 30 lower performing poor schools	3. 1st data collection in 60 school sample	5a. 2nd data collection in 60 school sample:

4. Select 8 case study schools - matched pairs	6. Leadership for literacy indictors → instrument to assess school leadership & management quality

5b. Analyse leadership

Quantitative data collection

In the quantitative component, fieldwork was conducted for one day in each of the 60 schools in February 2017 and again in October of the same year by a team of three fieldworkers (steps 3 and 5a in Figure 2). Data was collected on a battery of reading tests administered to students in Grades 3 and 6 and a number of instruments used to capture school characteristics, school climate, school functionality, teacher perceptions and leadership and management practices. These included structured interviews with the principal (P), deputy principal (DP), and two Grade 3 (G3) and two Grade 6 (G6) teachers; the administration of an anonymous self-administered educator survey to gauge perceptions; and learner book observations. Close-ended questions were preferred in the quantitative instrument development process. The reason relates to the broader study aim to develop a scalable instrument to measuring SLM where the critical issue in administering instruments at scale in the local context is generating low inference instruments given low levels of fieldworker capacity.

The object of gathering the test data twice in the same year was to compute student gain scores on the various literacy tests, and to link these to features of good school leadership with respect to promoting literacy instruction in the school, as established through the qualitative findings. An asset-based questionnaire was administered to one class of Grade 6 students in the February round of data collection to estimate the mean socio-economic status of each school's student composition.

Qualitative data collection

Qualitative data collection was done in the 8 case study schools in July, in between the two iterations of quantitative field work (step 5b in Figure 2). Each school was visited for three days by two experienced fieldworkers, during which time semi-structured, open-ended interviews were undertaken with the same educators listed above as well heads of

department[3] (HODs) for the intermediate phase[4] (IP) and foundation phase[5] (FP); textbooks and learner exercise books were inspected in the classes of the teachers interviewed; and the school library visited.

Crucial to the third leadership conundrum identified above, open-ended, probing interviews, combined with triangulation techniques – where the responses of one interviewee are tested for validity against the views of another interviewee on the same question – were employed in order to both to understand how leadership practices operate in schools and to penetrate the façade of 'socially acceptable' responses. In this regard, the case studies generated important descriptive findings of actual practices at the school level as distinct from reported practices.

Measuring the leadership for literacy resource domains

The three episodes of fieldwork (two quantitative and one qualitative) produced enormous quantities of data, which required aggregation. An intentionally 'blind' process was adopted in scoring schools along the four "leadership for literacy" dimensions from the quantitative and qualitative perspectives, respectively. Thus, quantitative scores emerging from applying the rubric measurement approach to the collected data were intentionally withheld from those analysing the case studies, using an independently developed set of rubrics, so as not to bias their rankings of schools based on the four resource dimensions. Discussion on the development of the two sets of rubrics and how each was used to inform the research questions follow.

Quantitative approach

Scoring rubrics are increasingly being used in economics to quantify competencies in areas of education management, assessment, or other systems technologies (Arcia et al., 2011; 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 policies compare to a described benchmark where the rubric descriptions guide the data to be collected. Our quantitative measurement approach centres on the development of a descriptive rubric to quantify competencies across the "leadership for literacy" theoretical dimensions which in turn can be distinguished into sub-dimensions as described in Table 2. The rubric development process involved mapping each resource dimension from the leadership for literacy framework into detailed descriptions of competence. For example, Table 3 provides descriptions for 5 elements under the human resources dimensions and how descriptions relate to quantitative scores of 1 (low) to 5 (high).

---

[3] HODs are mid-level school managers responsible for curriculum leadership in a subject or school phase.
[4] Grades 4-6
[5] Grades 1-3

Table 2: Sub-dimensions measure for Leadership for Literacy dimensions.

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

Table 3: 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) No respondents (0 of 3) identifies one or two specific people by name as being best at teaching reading in the school.	i) 1 of 3 respondents identifies one or two specific people by name as being best at teaching reading in the school.	i) 2 of 3 respondents identifies one or two specific people by name as being best at teaching reading in the school.	i) 2 of 3 respondents 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 students.	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	Less than a quarter of educator respondents teaching African or English language have either i) completed an Advanced Certificate in Education (ACE) or short	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	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	75% or more (but not all) educator respondents teaching African or English language have either i) completed an ACE or short course in teaching	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

expertise to teach reading in the school.	course in teaching language or reading OR ii) identify that English or African language was one their subject majors	ii) identify that English or African language was one their subject majors	ii) identify that English or African language was one their subject majors	language or reading OR ii) identify that English or African language was one their subject majors	language 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 the usual Integrated Quality Management System (IQMS) (a weak system-imposed evaluative mechanism).	School has a system for acknowledging its best teachers through rewards/awards other than the usual IQMS.	School has a system for acknowledging its best teachers through rewards/awards other than the usual IQMS. 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 the 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 the principal). The 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. 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. 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.

Having established definitions of competence the next step was to identify the type of close-ended questions that would be required to obtain enough information to determine if a school should be scored 1, 2, 3, 4 or 5. Close-ended questions rather than open-ended questions were administered, limiting high-level judgements required from fieldworkers. [6] Developing the close-ended questions was informed by the rubric descriptions. In the question or item-writing process the following questions guided us:

- Given the descriptions of competence required in the rubric, what type of data would we have to collect to objectively score each rubric element?
- Who would be the most appropriate respondent in a school to provide this data?
- What evidence-based information can we collect to verify respondents' answers to various SLM processes or practices?

This item writing process was iterative and various rounds of piloting of instruments were conducted in schools. Items relevant to scoring each school were incorporated into 6 sets of instruments that could be administered in a school over the course of a school day. Once data is collected and cleaned, a coding process is used to combine variables from 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 more subjective assessments of competence.

In total, over 500 variables were collected across the various instruments to generate 114 rubric elements which range from 1 (lowest possible score) to 5 (highest possible score). The elements vary in their construction using different types of data, namely; self-reported (respondent's recall of their experience or perceptions) and observational or evidence-based data. Almost half of the elements are coded using data that are triangulated in some way; for example, using responses from multiple individuals. The 6 leadership for literacy dimensions (with material resources split into time, availability of text and use of text) were obtained using a statistical procedure called principal components analysis to weight each index element in terms of the variation it explains in an underlying unobserved factor.

To assess the predictive validity of the leadership for literacy dimensions we use an education production function framework where Grade 6 literacy and reading outcomes are expressed as a function of specific 'leadership for literacy' index dimensions controlling for individual or home, and school characteristics – in particular school and student wealth. Outcome variables of interest included Grade 6 reading comprehension and vocabulary test results for over 2500 students, as well as oral reading fluency results in both English and

---

[6] The limited supply of high-quality fieldworkers (with strong educational backgrounds) 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.

African languages for roughly 600 Grade 6 students and 700 Grade 3 students. A value-added model was also estimated to determine whether 'leadership for literacy' indices explain any differences in literacy skills gained within a school year across the 60-school sample after accounting for student and school characteristics.

Qualitative approach

A rubric was constructed to collate the qualitative data on each of the 10 research questions listed in Table 1. A metric for each question was developed to assign a score to the schools with respect to that question. The method is illustrated with respect to research questions 1 and 2 (Table 4). The data relevant to these questions was made up of the responses to a question concerning the Grade level at which various literacy skills (knowing letter-sound relationships, reading words, reading isolated sentences, etc) should first be introduced to students[7]. The question was asked of four school leaders (principal, deputy principal, HODs for FP and IP), and four teachers (two Grade 3 teachers and two Grade 6 teachers of English). The four matched pairs of schools are designated letters A-D, while H and L indicate high- and low-performing schools.

Table 4: Assessment of case study schools regarding research questions Q1 and Q2

School	Research question	
	Q1: Are school leaders knowledgeable about teaching reading?	**Q2: Is this knowledge shared with educators across the school?**
A(H)	1	1
A(L)	1	1
B(H)	1	1
B(L)	1	1
C(H)	3	3 (E) 1 (Z)*
C(L)	1	1
D(H)	2	3
D(L)	1	1
Rubric	Assessment derived from the combined responses of 4 school leaders, as follows: 1 – No leaders show any knowledge of when to introduce elementary literacy skills 2 – At least the FP HOD shows signs of understanding, but this is incomplete 3 – At least the FP HOD and IP HOD exhibit a partial understanding, or the FP HOD exhibits an unambiguous understanding 4 – All school leaders exhibit an unambiguous understanding	Assessment derived from the combined responses of 4 leaders and 4 teachers, as follows: 1 – No convergence among educators 2 – Convergence but this is confined to individual teachers and is not shared by leaders 3 – Convergence among teachers and leaders within a grade or phase, or some convergence across the school, but this is incomplete 4 – Convergence between teachers and leaders across the school * There is convergence for English (Grade 6), but no convergence for isiZulu (Grade 3)

[7] This item was taken from the Progress in Reading Literacy Study (PIRLS) questionnaire.

Rubrics of this kind were constructed to assess the state of leadership in each of the case study schools on each of the 10 research questions (see Taylor and Hoadley, 2018 for details). The results of this analysis are shown in Table 5.

4. Key findings

Convergence across qualitative and quantitative findings

Weak leadership practices that are weakly associated with learning
Both the quantitative analysis of leadership practices in the full 60 school sample and the independent qualitative examination of the 8 case study schools revealed generally weak practices in all leadership for literacy domains. Where they did exist, these activities were inconsistent – if good leadership and management practices were discerned in the deployment of one type of resource, this was juxtaposed against weaknesses in how one or more of the other resources were deployed. These effects are 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 test). The best performing schools are no more likely to a have a larger percentage of the highest possible scores than the 10 worst performing schools.

Statistical multivariate analyses across the 60-school sample typically found little to no systematic relationship between most of the "Leadership for Literacy" dimensions and Grade 6 literacy or reading outcomes in English or African language in multivariate estimations controlling for a host of other school and student characteristics, including school wealth (see Wills and van der Berg, 2019). This result is not surprising, given that, where they exist in the sample, which is infrequent, better practices appear to be randomly distributed between and within schools. Where leadership practices are very weak and inconsistently applied, they can have little or no impact on test scores.

It should also be borne in mind that the sample of schools studied only included schools in challenging contexts rather than 'averaging' effects across different types of schools in which leadership may be more prominent and hence have a greater impact on learning. Despite the fact that the sampling strategy was designed to intentionally add as much performance variation into the sample that may possibly exist among these schools in three provinces (see

Figure 1), it was confined to a particular set of schools, namely those serving poor children in rural and township contexts. A recent analysis of linkages between measures of instructional leadership or school climate and Grade 9 mathematics outcomes using a nationally representative sample of no-fee public schools, was also unable to detect significant positive associations (Zuze and Juan, 2018).

These findings are replicated by the qualitative analysis, summarised in Table 5 which shows that not only are the ordinal scores of case study schools low to moderate on the

majority of 10 leadership indicators analysed, but also that better performing schools do not consistently score higher on every domain compared with their weaker performing counterparts (where H and L alongside the school designation A, B, C etc. indicate a higher or lower performer).

The consistency of findings concerning the very weak and inconsistent state of leadership from both qualitative and quantitative perspectives bolsters the face validity of the research findings, and the usefulness of the convergent parallel design is apparent.

Figure 3: Leadership for literacy scores across 114 rubric elements for the 6-best and 6-worst performing schools
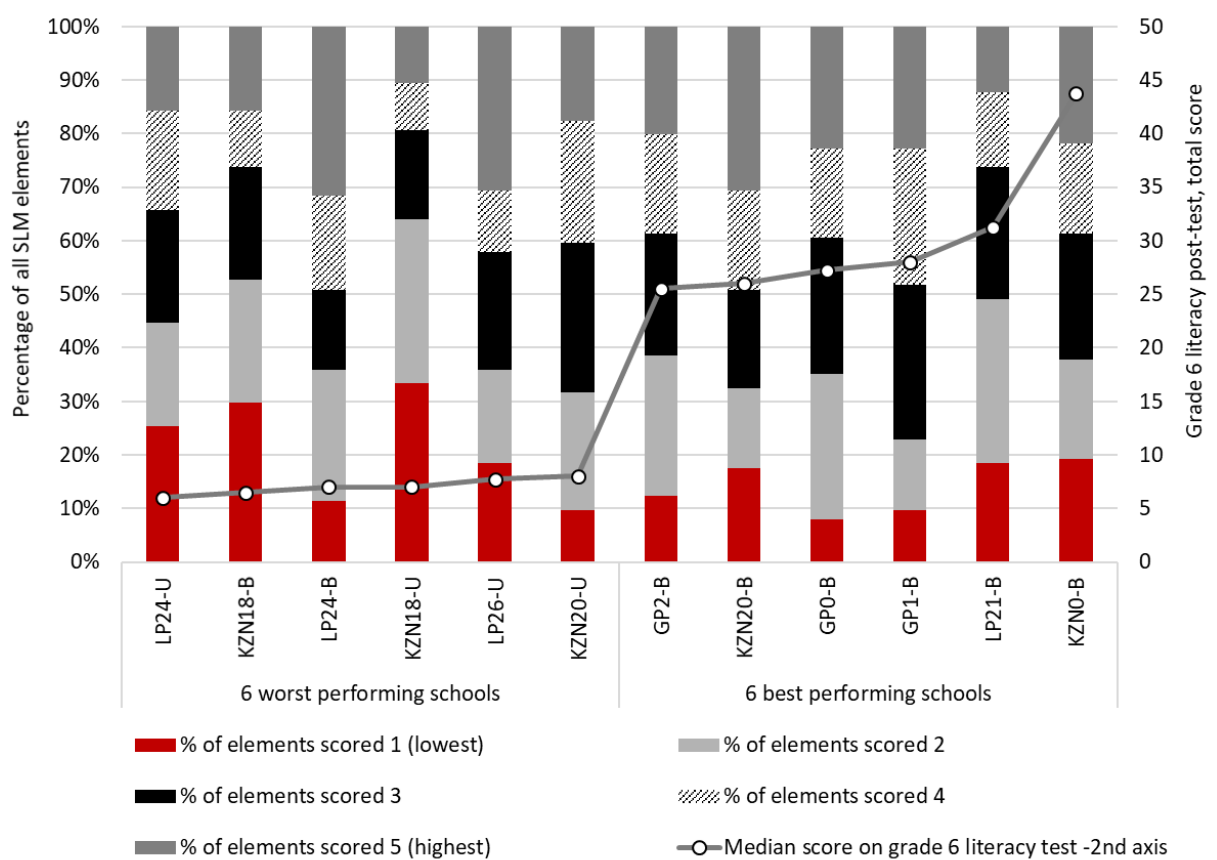
Source: Wills and van der Berg, 2019

Table 5: Assessment of case study schools on 10 aspects of leadership for literacy

School	Leadership resources									
	Knowledge			Human				Strategic	Material	
	Introduce literacy skills - knowledge exhibited	Introduce literacy skills - knowledge shared	Understand nature of reading pedagogy	Recruit staff according to expertise	Expertise in reading pedagogy utilised	Educators collaborate on pedagogical matters	External in-service training procured	Reading programme in place	Book spend according to policy	Library stocked and used in reading
A(H)	1	1	1	2	1	1	2	1	2	2
A(L)	1	1	2	1	1	1	2	1	4	2
B(H)	1	1	1	3	1	3	2	3	3	2
B(L)	1	1	2	2	2	3	2	4	2	1
C(H)	3	3/1*	2	3	1	3/1*	2	3	4	3
C(L)	1	1	2	3	3	3	2	3	4	2
D(H)	2	3	4	4	2	3	2	2	1	2
D(L)	1	1	2	1	2	1	2	1	1	2

* Relatively strong with respect to English; weak with respect to isiZulu

The value of human and knowledge resources

A further process in the qualitative analysis entailed reflection on the case studies in relation to the analytical framework, and especially, consideration of the relationship between different resources. The case studies suggested that the effective deployment and development of material, human and strategic resources is strongly mediated through the presence of knowledge resources, particularly those of incumbent leadership. The case studies reveal that strongly distinguishing higher performing schools in 2 of the 4 pairs of case study schools (pairs C and D) (see Table 5), is the presence of knowledge resources among school leaders. A key hypothesis emerging from the qualitative process is that if knowledge resources – the knowledge and understanding of reading and how it is best taught – provides the compass which enables school leaders to deploy the other resources at their disposal towards school-wide, effective reading instruction, then the most important vehicle for implementing this enterprise is the educator cohort at the school. Without willing and skilled teachers, the best books, libraries and reading programmes may create the illusion of good practice but lack the substantive engagement with young minds necessary to promote learning.

The quantitative analysis, could not detect strong relationships between literacy and knowledge resources across the 60 schools. This possibly is due to inadequate quantitative measures to establish the level of knowledge resources in the school. But we did find that schools with better human resource practices, experienced slightly higher gains in Grade 6 English literacy test scores. The linkages between the deployment and development of

human resources and learning, is supported through the strong positive association identified between the management and development of human resources and curriculum coverage in schools (as revealed in more evidence of work done in students' language workbooks or exercise books and by the extent to which educators reported that middle-managers checked their curriculum coverage). Many of the human resource elements that emerged as significant related to effective teacher selection practices by school governing bodies, hiring teachers with specialisms in language and teaching reading, teacher professional development, acknowledging excellence through systems of rewards and ensuring that there are enough leaders in the school to maintain systems of management.

In this respect, the quantitative and qualitative components lead to the same conclusions – school leaders should expend considerable effort in selecting, promoting, and deploying educators who exhibit the highest levels of motivation and expertise in reading pedagogy. While *de jure* government policy pays lip service to this ideal, the reality is very different. In half of the case study schools evidence for direct union interference in recruitment practices, or closed shop arrangements was detected, and may be happening in others where such evidence was not uncovered (Taylor and Hoadley, 2018). As described in Box 1, in one case the school started off parroting the official policy but probing soon revealed that the principal had almost no authority in making staff appointments. In another case, the principal was quite blunt about corrupt practices dictating appointments, when he said: … *[the union] always has the final word; money changes hands.*

Divergence across quantitative and qualitative findings

The agreement on the overall conclusions reached by the quantitative and qualitative analyses notwithstanding, the differences between the two sets of findings are also instructive. When drilling down to each of the leadership for literacy domains, the qualitative measurement results for the case study schools often contradicted the findings from the quantitative analysis. Table 6 provides examples of convergences and divergences between the two sets of analyses on 8 of the indicators shown in Table 5.

In only 2 of the 8 sub-dimensions of interest does there appear to be considerable convergence between the qualitative and quantitative findings. It is not surprising that results regarding the library converge – this is a low-inference, observable physical attribute. The other sub-dimensions require collecting self-reported recall, experiential or perception-based information for constructs or topics that cannot be directly observed, opening the door for the effects described above as *isomorphic mimicry* or respondents producing *socially acceptable responses*. Under these circumstances one is inclined to give more weight to the validity of the qualitative measures given that the open-ended, probing nature of the case study interviews, together with triangulation techniques are more likely to provide answers that are closer to what actually happens in schools, than the survey techniques which dominate quantitative studies. This not to say that qualitative methods are invariably, or even mostly, successful in this endeavour but the data offered below indicates that the probing and varied questioning techniques which characterise such research designs are better equipped to deal with the challenges of identifying misleading responses and getting

closer to the reality of behaviour of educators in schools. The case studies uncovered and addressed, through in-depth interviews, a number of instances of socially acceptable responses, of which one is described in Box 1.

Table 6: Qualitative and quantitative results Diverge (D), Diverge somewhat (DS) or Converge (C)

	Knowledge		Human				Strategic	Material
	Introduce literacy skills - knowledge exhibited	Under-stand nature of reading pedagogy	Recruit staff according to expertise	Expertise in reading pedagogy utilised	Educators collaborate on pedagogical matters	External in-service training procured	Reading programme in place	Library stocked and used in reading programme
Pair (A)	DS	C	C	C	DS	DS	DS	C
Pair (B)	D	DS	DS	DS	DS	C	D	C
Pair (C)	D	C	D	C	C	DS	DS	C
Pair (D)	C	C	D	D	D	DS	D	C
No. of pairs that converge	1/4	3/4	1/4	2/4	1/4	1/4	0/4	4/4

Box 1: Example of a misleading 'socially acceptable' response

In response to the question as to how educators are recruited for the school or promoted into leadership positions, the principal of C(L) started by providing a rational 'socially acceptable' response to the question, but probing revealed a very different reality. In her initial response the principal insisted that rigorous procedures are employed to recruit or promote educators who exhibit leadership qualities and curriculum expertise:

It's a good system because we select the best candidate, unlike other schools where excess teachers are deployed to schools. We are really working here, there is no time to waste and other teachers find it difficult to adjust when they come from schools which are not hard working like us.

Questioned about the role of the teachers' unions on selection committees, ostensibly as observers but frequently seen to be manipulating the process, the principal said that, while the views of union representatives are respected, the school selects educators according to merit:

We tell the union that we, the interviewing committee, will decide. [The] union is there to observe, not to select. We can take your opinion, but we make the decision. You must have a strong interviewing committee; you have to select people who are good, are educated and know the laws. We even select our teachers like this.

> However, on being probed, the Principal soon conceded that in reality these procedures are not followed entirely, if at all:
>
> *The department does try to deploy excess teachers and I have to take them. I'm not happy with that system: they come with a letter to say they must start at your school, so you don't have a choice. They deployed a high school teacher to a post in Grade 1. I have now put her in Grade 3, and I'm complaining to the Circuit Manager.*

This relative advantage of the qualitative method in getting closer to reality is compounded by the fact that the case studies were conducted by high level researchers. There was also more time in the schools for the case studies compared with quantitative process, and time allowed for probing of responses and increased in-depth questioning. In contrast, conducting quantitative studies in relatively large samples of schools, in order to provide for statistically valid results, determine that the time spent in each school be kept to a minimum and that low-cost fieldworkers be employed. Quantitative studies deliberately reduce fieldworker interpretation, through highly structured instruments, in order to improve the reliability of the data. But this reduces the ability of fieldworkers to detect misleading responses, in turn reducing the validity of the response. Qualitative findings on the other hand, although getting closer to identifying the practices actually occurring in schools, are not generalizable because of the small sample size and relatively more subjective nature. This situation recalls Einstein's (1921: 1) apparently paradoxical statement:

> *As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.*

**Incorporating triangulation into quantitative data collection**

Incorporating triangulation into the qualitative research component was vital for probing and uncovering overall leadership realities in schools. In the spirit of triangulation, an interesting addition to the quantitative data collection process was the use of a self-reported survey instrument administered to all teachers in the school. This tool revealed stark differences in teachers' experiences and interaction with school management teams within the same schools. For example, Figure 4 reports the percentage of teachers in the 5 best and 5 worst performing schools across the 60 school study, identifying specific frequencies with which their head of department (HoD) – a middle manager in a school - checks to see how much of the curriculum they have taught. Teacher's experiences evidently vary within the same schools. Some educators report 'never', others report 'weekly' checks or multiple checks during the week.

This highlights that drawing research conclusions from the quantitative data is strongly dependent on who is interviewed in the school environment.[8] Incorporating validation and a wider respondent base into data collection is necessary and could contribute to more

---

[8] Yet, international school surveys such as PIRLS or TIMSS will gather school climate or leadership related indicators from just one or two teachers in a school.

reliable data. Nevertheless, this was still not sufficient to match the validity of the case study process or overcome the need for high level researchers.

Figure 4: Teacher responses in schools - "How often does your Head of Department in this school check to see how much of the curriculum you have taught?"

How often does your HoD in this school check to see how much of the curriculum you have taught?
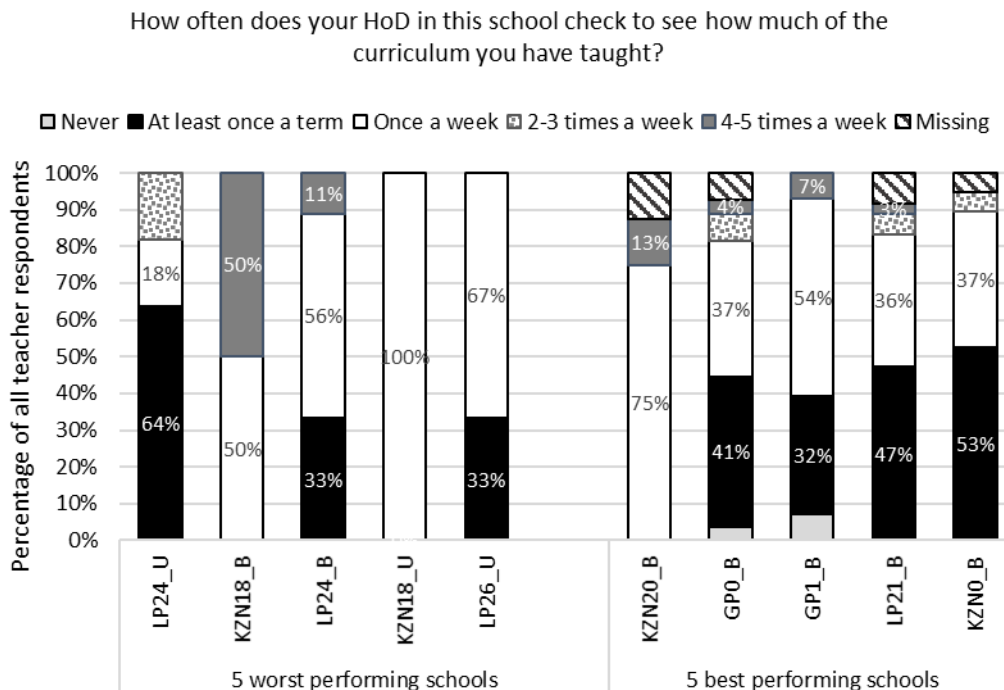
5. Conclusion

In effecting a mixed methods design in relation to the question of school leadership for literacy we attempted to address some of the central methodological difficulties implicated in producing the leadership conundrum – what we described as contradictory findings observed across qualitative and quantitative research in educational leadership and management. In concluding the paper, we reflect on the extent to which the study addressed three central difficulties we identified earlier, in relation to theory, the perennial problems of validity and reliability, and the difficulty of selecting an appropriate sample of schools.

Theory

A good theory is essential to ensuring that we are measuring the right things. From a Popperian perspective, no theory is ever complete and is always subject to refutation or modification. The theory we developed from an exhaustive literature review proved to be useful in the systematic search for data to illuminate the research question, structuring the analysis of the data and providing insights into the behaviour of school leaders. These insights, in turn, suggested that, not only are the four kinds of resources identified in the theory essential to the development of successful reading pedagogy across the school, but

that they exist in a hierarchical relationship with one another. Thus, knowledge resources on the part of school leaders are prerequisite to selecting, deploying and supporting the human resources able to effectively teach reading and writing; expert teachers, in turn, are key to the formulation and implementation of a school-wide reading programme; which is dependent on the effective use of time and high-quality reading material.

The multi-dimensional theoretical framework was employed to guide data collection for both the qualitative and quantitative elements. Although it only focuses on educational leadership from the viewpoint of supporting literacy development, the framework incorporates a wide range of dimensions that we discovered were differentially amenable to measurement across quantitative and qualitative techniques. Contrary to Robinson et al's (2008) view that narrow frameworks leave effects under-detected, we argue that future research on measuring leadership and management would be supported by focusing and measuring a few things well. The importance of finding ways of measuring knowledge resources in large samples should be of particular interest. We hypothesise that this may be where the residual in relation to weak findings of school management studies generally might lie.

Sampling

The sampling strategy for the study was designed to intentionally add as much performance variation into the sample for both the quantitative and qualitative arms. However, our research question confined us to schools serving poor children in rural and township contexts. What we found, both in seeking better performing schools and in the data we generated, were remarkable levels of similarity across schools. The matched pairs methodology has been attempted a number of times in South Africa, without unqualified success (DPME/DBE, 2017; Hoadley and Galant, 2015; Taylor et al., 2013), and the difficulty lies both in identifying schools that are performing sufficiently above expectations given their demographics to constitute true 'outliers' and in the uniformity of schooling in these contexts.

Validity and reliability

Once we had identified the right things to measure through our theory, we had to decide how best to measure them. Both quantitative and qualitative approaches have in-built structural deficiencies (i.e. these deficiencies are not due to inadequate application but come with the territory and won't go away), rendering either, on its own, inadequate to the task. This is especially so in relation to the issues of validity and reliability:

- The quantitative approach has a *validity problem* – is the data reflective of reality? – because of the prevalence of the effects of isomorphic mimicry, the production of socially acceptable responses and the limitation of fieldworker interpretation. This deficiency effectively nullifies the third and fourth aims of the study, which were to develop a scalable SLM instrument with predictive validity. On the positive side, quantitative methods are superior in their detection of 'average' effects over a statistically significant sample, and hence of generalisation.

- The qualitative perspective has a *reliability problem* because of the small sample and the play of subjectivity in collecting and interpreting the data. Will we come to different conclusions if we use a different sample, use different fieldworkers or interview different respondents in the school? On the positive side, qualitative methods are more likely to uncover some of what is actually happening in schools.

This study makes an important contribution in highlighting the importance of mobilising the advantages of both approaches to unravelling the skein of compliance, very prevalent in management practices in highly bureaucratised systems. Quantitative measurement using self-reported and interview style assessments will be limited in their ability to capture real behaviours, activities and processes until more attention is given to innovative approaches to overcome these biases. Improved measurement would benefit from closer interdisciplinary collaboration between educationists, economists, anthropologists, and psychometricians.

Overall, while the qualitative and quantitative findings confirmed that effects were weak and inconsistent, the nature and extent of those effects differed across the survey and case studies. What the study highlighted were the difficulties entailed in developing a scalable instrument to measure school leadership and management in challenging contexts, especially where fieldworker expertise is limited (a common issue across developing contexts). Paying careful attention to issues of sampling, the development of theory and social desirability bias, we were in some ways able to generate more robust findings. But it was clear that different methodologies were able to pick up different aspects of SLM and show effects, and at times these findings did not reinforce each other. The question remains regarding the extent to which we leave high inference, penetrative research to qualitative work, or attempt to render surveys more high inference? The latter does, however, have significant cost implications for SLM research given the seeming necessity for high level fieldworkers and in-depth interview to generate robust responses. Alternatively, we need to think through more complex designs where we mix and match design components in a way that offers the best chance of answering our specific research questions. The present study is one such attempt, and while each perspective does not eliminate the weaknesses of its counterpart, putting the findings of the two together provide far more valuable insights than are produced by each on its own.

6. References

Arcia G, Macdonald K, Patrinos HA, et al. (2011) *School Autonomy and Accountability*. *System Assessment and Benchmarking for Education Results*. System Assessment and Benchmarking for Education Results 94444. Washington, D.C. DOI: 10.1787/5k9h362kcx9w-en.

Bloom N and Van Reenen J (2010) New approaches to surveying organizations. *American Economic Review* 100(2): 105–109. DOI: 10.1257/aer.100.2.105.

Branch GF, Hanushek EA and Rivkin SG (2012) *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*. *National Bureau of Economic Research Working Paper Series*. 17803. Cambridge, MA. DOI: 10.3386/w17803.

Bush T and Heystek J (2006) School leadership and management in South Africa: principal's perceptions. *International Studies in Educational Administration* 34(2): 63–76. Available at: http://www.emeraldinsight.com/doi/10.1108/IJEM-07-2014-0101.

Caracelli VJ and Greene JC (1997) Crafting mixed-method evaluation designs. *New Directions for Evaluation* (74): 19–32. DOI: 10.1002/ev.1069.

Creswell JW (2003) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications, Inc.

Crouch L and 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. *Social Dynamics* 27(1): 60–78. DOI: 10.1080/02533950108458704.

Deaton A (2010) Instruments, Randomization, and Learning about Development. *Journal of Economic Literature* 48(2): 424–455. DOI: 10.1257/jel.48.2.424.

Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine* 210: 2–21. DOI: 10.1016/j.socscimed.2017.12.005.

DPME/DBE (2017) *Implementation Evaluation of the National Curriculum Statement Grade R to 12 Focusing on the Curriculum and Assessment Policy Statements (CAPS)*. Pretoria.

Edmonds R (1979) Effective schools for the urban poor. *Educational Leadership* 37: 15–24.

Einstein A (1921) Geometry and experience. *Sidelights on relativity*.

Fleisch B and Schöer V (2012) Large-scale instructional reform in the Global South: Insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education* 34(3): 1–12.

Fleisch B, Schöer V, Roberts G, et al. (2016) System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics. *International Journal of Educational Development* 49: 157–174.

Grissom JA, Kalogrides D and Loeb S (2015) Using Student Test Scores to Measure Principal Performance. *Educational Evaluation and Policy Analysis* 37(1): 3–28.

Hallinger P and Heck RH (1996) The Principal's Role in School Effectiveness: An Assessment of Methodological Progress, 1980-1995. In: Leithwood K (ed.) *The International Handbook of Research on Educational Leadership and Administration*. New York: Kluwer Press, pp. 1980–1995.

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

Hoadley U and Galant J (2015) The organisation of schools succeeding against the odds. *South African Review of Education* 12(2): 29–52.

Hoadley U, Christie P and Ward C (2009) Managing to learn: Instructional leadership in South African secondary schools. *School Leadership* 29(4): 373–389. DOI: 10.1080/13632430903152054.

Jang EE, McDougall DE, Pollon D, et al. (2008) Integrative mixed methods data analytic strategies in research on school success in challenging circumstances. *Journal of Mixed Methods Research* 2(3). DOI: 10.1177/1558689808315323.

Kotze J, Fleisch B and Taylor S (2018) Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa. *International Journal of Educational Development*. DOI: 10.1016/j.ijedudev.2018.09.004.

Leithwood K, Louis KS, Anderson S, et al. (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 and Scur D (2017) *Developing Management: An expanded evaluation tool for developing countries*. RISE working paper series 16/007. Available at: http://www.riseprogramme.org/content/rise-working-paper-16007-developing-management-expanded-evaluation-tool-developing-countries.

Maden M (2001) *Success Against the Odds, Five Years On: Revisiting Effective Schools in Disadvantaged Areas*. London: Routledge Falmer.

Mertler C (2019) *Mertler, C. (2019). Introduction to Educational Research*. Los Angeles: SAGE Publications, Inc.

Mullis I, Martin M, Foy P, et al. (2017) *PIRLS 2016: International Results in Online Informational Reading*. International Association for the Evaluation of Educational Achievement (IEA). Available at: http://timssandpirls.bc.edu/pirls2016/international-results/.

Onwuegbuzie AJ and Johnson RB (2004) Mixed Research. In: Johnson RB and Christensen LB (eds) *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. 2nd ed. Needham Heights, MA: Allyn & Bacon, pp. 408–431.

Piper B (2009) *Integrated Education Program Impact Study of SMRS using Early Grade Reading Assessment in three provinces in South Africa*. Research Triangle Park, NC. Research Triangle International.

Pritchett L, Woolcock M and Andrews M (2013) Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation. *The Journal of Development Studies* 49(1): 1–18.

Robinson VMJ, Lloyd CA and Rowe KJ (2008) The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly* 44(5): 635–674. DOI: 10.1177/0013161X08321509.

Tashakkori A and Creswell JW (2007) Editorial: Exploring the Nature of Research Questions in Mixed Methods Research. *Journal of Mixed Methods Research* 1(2): 207–211. DOI: 10.1177/1558689807302814.

Tashakkori A and Teddlie C (1998) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, CA.: SAGE Publications, Inc.

Tashakkori A and Teddlie C (2003) *Handbook of Mixed Methods in Social and Behavioral Research*. Tashakkori A and Teddlie C (eds). Thousand Oaks, CA: SAGE Publications, Inc.

Taylor N and Hoadley UK (2018) *Leadership for Literacy: Exploring exceptional leadership practices in township and rural primary schools serving poor communities in South Africa. Report on the Case Study Schools*. Research on Socio-Economic Policy. Stellenbosch.

Taylor N, Gamble J, Spies M, et al. (2013) Chapter 4: School leadership and management. In: Taylor N, van der Berg S, and Mabogoane T (eds) *Creating Effective Schools: Report of South Africa's National School Effectiveness Study*. Cape Town: Pearson Education.

Taylor S, Cilliers J, Prinsloo C, et al. (2017) *The Early Grade Reading Study: Impact evaluation after two years of interventions. Technical Report. Unpublished*. Pretoria.

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

Wills G and van der Berg, S. (2019) *Measuring leadership and management and its linkages with literacy in rural and township primary schools in South Africa*. Research on Socio-Economic Policy. Stellenbosch.

Witziers B, Bosker RJ and Krüger ML (2003) Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly* 39(3): 398–425. DOI: 10.1177/0013161X03253411.

Zuze T and Juan A (2018) *School Leadership and Local Learning Contexts in South Africa*. Research on Socio-Economic Policy. Stellenbosch.

7. Acknowledgements

8. Funding