Great Teaching Toolkit

School Environment & Leadership: Evidence Review

Rob **Coe**

**Evidence that
school leadership and
environment matter**

In partnership with

**Cambridge Assessment
International Education**

The suggested citation for this document is

Coe, R. (2022). Evidence that school leadership and environment matter (School environment and leadership: Evidence review). Evidence Based Education.

https://evidencebased.education/school-environment-and-leadership-evidence-review/

IGNIO

Innovation by design

Contents

Introduction

This paper is part of a series of four that together comprise the *Great Teaching Toolkit School Environment and Leadership Evidence Review*. In order to cater for different audiences, we have split the findings from our evidence review into four separate, but inter-related, documents, of which this is the third.

The first, written primarily for practitioners, and intended to have a constructive, action focus, sets out our *Model for School Environment and Leadership*–the school-level factors that can inhibit or enhance the classroom interactions that promote effective learning. The second explains in technical detail the key methodological problems faced by research in school leadership, and hence why we are sceptical of many of its claims. The third (this document) identifies a selection of studies that we believe contain the most defensible claims and the strongest evidence about the factors we have included in the *Model for School Environment and Leadership*. The fourth provides technical details of the literature search and synthesis process that underpins the other three.

You can find links to download all four papers **here**.

Evidence that school leadership and environment matter

Despite the limitations of much of the available evidence (see paper 2), there is undoubtedly also good research within this field. For example, Robinson et al.'s (2009) Best Evidence Synthesis presents a comprehensive picture of the state of the evidence at that time. A more recent review by Grissom et al. (2021) provides new and compelling evidence about the impact of principals on student attainment, summarises—with appropriate caveats—what we know about effective leadership behaviours and the overall state of the literature in this field. The work of Bryk and colleagues (A. Bryk et al., 2012; A. S. Bryk, 2010; A. S. Bryk et al., 2010) demonstrates that key features of a school's environment and culture are predictive of future growth in student attainment. A key paper by Kraft and Papay (2014) and other related work (e.g., Johnson et al., 2012) demonstrates that 'professional environment' predicts both the growth in measures of the individual effectiveness of teachers and their likelihood of staying in post. The Dynamic Model of Creemers and Kyriakides (2011) provides a model of school effectiveness and improvement based on strong theory and robust empirical testing.

This paper summarises our best knowledge from these and other studies in the form of five claims about school leadership and environment that we believe are both scientifically meaningful and supported by the best current evidence. A sixth claim is that there are some key things we still do not know.

In summary the five claims are:

Claim 1: Individual school headteachers may have significantly different outcomes.

Claim 2: Retention of both teachers and headteachers matters.

Claim 3: Individual teacher effects are only part of the story.

Claim 4: Some school-level characteristics are associated with student outcomes.

Claim 5: Some school-level characteristics predict school and teacher growth.

The final (6th) claim is that we still do not know:

- The malleable, well-defined behaviours of headteachers that make a difference.
- The size and nature of the contribution of school leaders who are not headteachers (e.g., assistant heads).
- How to train and support headteachers and school leaders to be more effective.
- Anything about the size and distribution of teacher or headteacher effects from different contexts (e.g., UK).
- The extent to which teacher or headteacher effects depend on interactions with key factors.

The selection of studies for this review came from a systematic process of identifying relevant studies, and extracting and synthesising their main claims and methods. A full description of the methods used in the search, screening, extraction and synthesis processes can be found in paper 4.

Claim 1: Individual school headteachers may have significantly different outcomes

There are now a number of studies that have longitudinal student attainment data collected over a period of years from the same school with multiple principals, and sometimes matching the same principal across multiple schools. By estimating separate 'effects'[1] for the principal and school it is possible to distinguish the components of student attainment, after controlling for explanatory variables such as prior attainment and social mix, that are associated with an individual principal or school. These models typically estimate the principal 'effect' as the difference between the attainment under that principal and attainment in the same school under all other principals. Grissom et al. (2021) identified six of these studies conducted in the US since 2000.[2] In addition, there are at least two similar studies conducted in Canada, reviewed in a meta-analysis by Liebowitz and Porter (2019).[3]

Of course, we should remember that principals are not randomly assigned to schools, so any unobserved factors that are associated with those choices (or allocations) could explain these 'effects'. For example, if some principals are able to judge that a school is about to improve and move there just as results begin to rise, we might attribute this improvement to their impact. However, it is hard to come up with plausible confounding interactions of this type. In this case, therefore, it probably is appropriate to interpret the coefficients from the regression models as causal effects of the principal on student attainment.

Grissom et al. (2021) report that "a 1 standard deviation increase in principal effectiveness increases the typical student's achievement by 0.13 standard deviations in math and 0.09 standard deviations in reading". Liebowitz and Porter's (2019) estimate is comparable: 1 SD in principal effectiveness corresponds to between 0.05 and 0.41 SD in student learning gains. These associations are "nearly as large" as most estimates of the effects of individual classroom teachers on student attainment (Grissom et al., 2021), but the principal may have this effect across all students in the school, in addition to any effects of individual teachers. If these results are replicated (ideally, with data from multiple principals in the same school and each principal observed in multiple schools, and from a wider range of contexts) it would certainly be appropriate to conclude that individual school principals make an important difference to the outcomes of their students.

---

1   The word *effects* is conventionally used to refer to the coefficients in a regression model, even where there is no explicit argument about why a causal interpretation is appropriate, and such an interpretation may not even be intended. This convention is unfortunate, but well established. In an attempt to avoid confusion, We will use inverted commas around the word 'effects' unless a causal claim is intended.
2   i.e, Bartanen, 2020; Branch et al., 2012; Chiang et al., 2016; Dhuey & Smith, 2018; Grissom et al., 2015; Laing et al., 2016
3   i.e., Coelli & Green, 2012; Dhuey & Smith, 2014

However, in most of these studies, the number of principals observed at each school is quite low (typically two or three) and even where principals can be tracked across schools, the average number of schools observed with the same principal is even lower; it takes a lot of schools and many years of data to overcome these limitations. As a result, the separation between school and principal effects may not be as clean as it would be with more extensive data; some of each may still be confounded with the other. This challenge is strengthened by evidence from Bartanen et al. (2022), who make use of three large US datasets (13 continuous years of student data from all schools in Tennessee and Oregon; 19 years of data from New York City). Adopting a different analytical approach from previous studies, they compare value-added estimates of the performance of a school in successive years, split by whether the principal is the same or different. In all cases, correlations decrease as the number of years' gap increases, but they find no difference between year-year correlations with the same principal in post and those with a different principal. It is hard to reconcile the claim that principals make a difference to student outcomes with the finding that inter-year school performance is no more similar under the same principal than under a different principal. This result may not be the final word on the impact of school principals, but it does suggest that the estimates from the previous studies reviewed by Grissom et al. (2021) are probably an upper bound on their true impact, and the truth may well be closer to Bartanen et al.'s (2022) estimate of zero impact.

Claim 2: Retention of both teachers and headteachers matters

Among the same studies cited in the previous section are some other important findings. One is that a change of principal is associated with a dip in immediate school performance. Although the effect is not large ("0.01 to 0.02 standard deviations for the average student in the school"; Grissom et al., 2021, p. 51), it indicates a net cost associated with principal turnover. The analysis by Coelli and Green (2012) is distinctive in that they modelled the change in the impact of a principal over time, starting with a much smaller impact in their first year and gradually converging towards the principal's 'true' effect. This dynamic model is symmetric in the sense that weaker principals gradually have increasingly negative effects. Hence it is not modelling principals' growth and improvement in the job, but simply the lag in time before their impact can be observed. Coelli and Green's analysis suggests that this dynamic model, in which impact takes time to take effect, is a better fit to their data. According to the dynamic model of increasing effects, after five years in post, a typical principal has reached only 75% of the impact they would ultimately achieve. After three years, barely more than half their potential impact is observable. Unfortunately for the US school system, however, Grissom et al. (2021) report that the average principal's tenure in a school in the US is only four years, and at any given time around 30% have three or fewer years total experience as a principal; figures on tenure and experience have been in steady decline in the US since 1988 and are even worse for high-poverty schools.

Grissom et al. report the evidence from correlational studies that "more experience as a principal is associated with more positive outcomes" (2021, p. 50). Clearly, this association may be explained in a number of ways, among them the implied interpretation that, on average, principals grow in the job and become more effective with more experience. While there may be other reasons to think this is likely—most kinds of expertise take a long time to develop (Ericsson, 2009)—it is also likely that schools seek to retain effective principals and that less effective principals may be more likely to leave. Hence the true effect of principal experience on student attainment may be smaller than the correlations suggest.

Branch et al. (2009, 2012, 2013) report an interesting analysis of data from elementary and middle schools in Texas between 1995-2001 that includes over 7,000 principals. They use value-added models with school fixed effects, that "compare average student achievement gains in the same school under different principals" (2013, p. 3), and match this with data on teacher turnover. They find that the variation in principal quality is larger in high-poverty schools.

They also find a relationship between the quality of a principal and which teachers leave: "less-effective teachers are more likely to leave schools run by highly effective principals" (2013, p. 5). This suggests that one of the ways principals affect outcomes is through differential teacher retention.

Branch et al. report that principal turnover is higher at lower-achieving and higher-poverty schools. Less effective principals are more likely to leave both their current school and the public school system, but the patterns are somewhat mixed. For example, only in low-poverty schools does principal quality consistently predict likelihood of leaving: in these more advantaged schools, 76% of top-quartile principals remain in the same school after their third year, compared with only 59% of those in the bottom quartile. In high-poverty schools the differences are smaller, and it is principals in the middle range of quality who are most likely to remain: both low- and high-performers are slightly more likely to move on—perhaps for different reasons, though we have no direct evidence about this. Furthermore, leaving their current school often means moving to another school. If they move, low-performing principals in high-poverty schools are more likely to move to another school than low performers in more affluent schools, who are more likely to leave the system. Branch et al. (2013) speculate this may be because poor performance in high-poverty schools is harder to identify (principals are given the benefit of the doubt in subsequent job applications), or that district administrators have to bargain, with an offer of a job in another school, to get them to move.

Although there is considerable evidence that teacher turnover is higher in high-poverty schools in the US (e.g., Hanushek et al., 2004), more recent analyses indicate that "the high rates of teacher turnover observed in such schools are largely explained by the poor working conditions in these schools, rather than the students they serve" (Papay & Kraft, 2017, p. 18). A range of outcomes, including school-level teacher turnover and job satisfaction, student attainment and academic growth, are predicted by working conditions:

> While the elements commonly thought of as working conditions—such as planning time, school facilities, or instructional resources—are important, the elements that are social in nature tend to matter the most. These include (1) collegial relationships, or the extent to which teachers report having productive working relationships with their colleagues; (2) the principal's leadership, or the extent to which teachers report that their school leaders are supportive and create school environments conducive to learning; and (3) school culture, or the extent to which school environments are characterized by mutual trust, respect, openness, and commitment to student achievement. (Johnson et al., 2012, p. 25)

Daly et al. (2017) document the various ways in which leadership churn negatively impacts schools, and many of these factors apply equally to teacher turnover (Ronfeldt et al., 2013). Among the costs of turnover we might list:

- Disruption of social relationships and networks, and of the collaboration, cohesion, trust and support that depend on them;

- Disruption of curriculum and instructional cohesion, as new recruits adjust, learn and fit in;

- Loss of institutional knowledge and organisational memory;

- The costs of making any changes in practice or policy (often associated with new leaders);

- Having to restart the clock on the time required for changes to take effect and embed, as the benefits of any changes are often lagged;

- Costs of training and preparation for new staff; and

- Costs of recruitment and induction for new staff, including time spent mentoring and supporting.

If it is right that principals typically need five years to have a positive impact (Seashore Louis et al., 2010, p. 164) but the average tenure of a school principal is only four years (Grissom et al., 2021, p. 25), then it may be fair to describe a common state in schools as being in constant flux: continually bearing the cost and disruption of making changes and building new relationships, while never reaping the rewards those changes might have brought before the next round of change begins. Such a state will be both frustrating and stressful for all involved, undermine productivity and contribute to further churn, creating a vicious cycle of decline.

Of course, there may also be benefits to staff turnover, for example, if new staff bring new ideas, if recruitment offers the chance to better match job requirements to staff capabilities, or if the process simply results in higher net quality of staff. In fact, several studies (reviewed in Ronfeldt et al., 2013) have found that teachers who leave a school are on average less effective than those who stay—though it may also be that new joiners are also less effective than the average.

Empirical evidence that there is a negative effect of teacher turnover, over and above the direct impact of any change in net quality, is presented by Ronfeldt et al. (2013). They analyse data from grades 4 and 5 in New York, analysing by grade within school, as they (mostly) do not have individual teacher-student allocations, controlling for prior achievement and other covariates. The relationship between turnover and learning gains is clear, though not large: estimates of the difference in value-added between a school-grade-year in the bottom quartile of turnover (who had zero turnover) and those in the top quartile (who lost 37% of their staff, on average) are a reduction of around 2-4% of a standard deviation in test scores in both English and mathematics. These coefficients are slightly larger in schools with lower attainment or higher proportions of minority students.

Claim 3: Individual teacher effects are only part of the story

While a substantial body of research on teacher effects has established that individual teachers make a difference to student outcomes (e.g., Chetty et al., 2014), it is also clear that the impact a teacher has on a given class, in a given school, in a given system, is more than just a constant characteristic of that teacher. In most estimation models, any differences in the value-added achieved by different classes taught by the same teacher are treated as noise: random error in the estimation of a teacher effect. Analyses in which the same teacher is observed in more than one school are so rare that we know little about how transferable these 'effects' are.[4]

And yet, the evidence cited above—that an additional effect can be attributed to the school principal— along with a range of associations between school-level characteristics and learning gains, suggests that how effective a teacher is depends on more than just that teacher. The environment in which they work, whether determined by the actions of school leaders, their fellow teachers, other staff, students, parents, wider community and context, makes a substantial difference to how much students learn—over and above anything an individual teacher does.

It seems likely that even the best teacher, if put into an unsupportive environment, will not be successful:

> "For teachers who have the misfortune of trying to deliver high-quality instruction and improve their craft amidst organizational dysfunction, continually reshuffling the staff in search of teachers who can be successful in spite of organization limitations is likely be an ineffective strategy to improving instructional quality. To systematically improve student performance, school and district leaders need robust evidence about the strengths and weaknesses of both individual teachers and the school organization as a whole." (Kraft et al., 2016)

An example of teacher peer-effects comes from a study by Jackson and Bruegmann (2009), with data from grades 3 to 5 in North Carolina. They found that the same teacher appears more effective when their colleagues teaching the same year group are more effective. These 'spillover' effects are greatest for less-experienced teachers and, in part at least, are retained in subsequent years, suggesting that learning from peers may account for the effect.

---

4   But see Chetty et al. (2014) for an empirical check on the assumption that teacher effects are stable across movements between schools.

Kraft and Papay's (2014) study of the growth in effectiveness of teachers working in schools with different professional environments shows that more supportive contexts produce substantially bigger improvements over time. Their data come from teacher, student and state mathematics test records from schools in North Carolina, from 2000-2010, for students in grades 3 to 8. They find that, on average, teachers show rapid improvements in their effectiveness over the first three years of teaching experience, then continue to improve but much more slowly: a typical teacher improves almost five times as much in their first three years as they do in the next six. However, individual trajectories show considerable variation; while some teachers achieve four times the average growth, others actually decline over time. An important implication of this is that the effectiveness of an individual teacher is not constant: a weak teacher today can grow into a strong teacher in a few years; an initially effective teacher may plateau or decline and be overtaken by others who were previously less good.

Claim 4: Some school-level characteristics are associated with student outcomes

Many, many studies have looked for and reported associations between school-level characteristics and student outcomes. As discussed in *Methodological challenges in school leadership research*, the quality of these studies and the credibility of their claims is quite variable. Among them, we have identified four specific pieces of work that stand out as being of high quality, comprehensive and providing a clear narrative that may be useful to school leaders.[5] The first is a systematic 'Best Evidence Synthesis' by Robinson et al. (2009). The second is a systematic (and generally more critical than many) review by Liebowitz and Porter (2019). The third is another systematic review, by Grissom et al. (2021). The fourth is a strand of work by Creemers and Kyriakides (e.g., Creemers & Kyriakides, 2011; Kyriakides et al., 2019) on their 'Dynamic Approaches' to educational effectiveness and improvement.

Robinson et al. (2009)

For many years, Robinson et al.'s (2009), systematic review of the leadership literature was the most robust and authoritative guide to the best available evidence. The *Best Evidence Synthesis* approach combines comprehensive and systematic searching with a transparent and rigorous review process that focuses on practical implications for school leaders. It limits its focus to studies that measure student outcomes and a range of aspects of leadership. Robinson et al.'s review makes good use of theory and triangulation of results using mixed methods. They also provide helpful conceptualisations of otherwise fuzzy concepts, such as 'transformational leadership' (2009, p. 86).

Although Robinson et al. talk about "the impact of leadership on student outcomes" and "the dimensions of school leadership that make a difference to students", the vast majority of the evidence included is correlational, so words like *impact, effects* or *make a difference* are somewhat misleading. They do acknowledge that causation is complex, but then go on to use this kind of causal language unproblematically. This is unfortunate, because knowing (a) that certain behaviours of leaders are associated with better student outcomes is not the same as knowing (b) that these behaviours can be changed and (c) that such change interventions produce measurable improvements in valued outcomes. In fact, (b) and (c) may not follow at all from (a), and, even if they do, the strength of the relationship will almost certainly be very different.

5   Of course, there may well be other good studies and reviews out there; these are illustrative examples.

A summary of the results of the meta-analysis of different leadership dimensions is that five dimensions have the strongest associations with attainment (Robinson et al., 2009, p. 95, Table 6):

1. Establishing goals and expectations. Includes the setting, communicating, and monitoring of learning goals, standards, and expectations and the involvement of staff and others in the process so that there is clarity and consensus about goals. ES = .42 (.07) (49 effect sizes from 7 studies).

2. Resourcing strategically. Involves aligning resource selection and allocation to priority teaching goals. Includes provision of appropriate expertise through staff recruitment. ES = .31 (.10) (11 effect sizes from 7 studies).

3. Planning, coordinating, and evaluating teaching and the curriculum. Direct involvement in the support and evaluation of teaching through regular classroom visits and the provision of formative and summative feedback to teachers. Direct oversight of curriculum through school-wide coordination across classes and year levels and alignment to school goals. ES = .42 (.06) (80 effect sizes from 9 studies).

4. Promoting and participating in teacher learning and development. Leadership that not only promotes but directly participates with teachers in formal or informal professional learning. ES = .84 (.14) (17 effect sizes from 6 studies).

5. Ensuring an orderly and supportive environment. Protecting time for teaching and learning by reducing external pressures and interruptions and establishing an orderly and supportive environment, both inside and outside classrooms. ES = .27 (.09) (42 effect sizes from 8 studies).

Note that, although the language use is of 'effect size' and 'effects', all the evidence on which the claims are based is correlational.

A second strand of work in Robinson et al. (Robinson et al., 2009, Chapter 6) uses evidence from evaluations of interventions. These were taken from either a systematic review of teacher professional learning or from other interventions evaluated in New Zealand schools. As these are intervention studies, they may remove some of our concerns about the confusion of correlation with causation in the meta-analysis cited above. However, separating the role of leadership and its precise mechanisms in mediating the effects of these interventions was often not explicitly part of the design of those studies, and generally has to be inferred somewhat approximately. The fact that there is convergence across the two strands can indeed be taken as evidence that these dimensions may well be important.

Knowing precisely how important, and in what ways they can be leveraged in practice to create improvements, seems less clear, however.

A third strand of evidence (Robinson et al., 2009, Chapter 7) draws on a systematic review of school-home-community interactions. A number of interventions to engage parents and community in the work of the school are reviewed and presented as supporting the case that school leaders should prioritise these kinds of activities.

Perhaps the most practically useful section of Robinson et al.'s review is where they break down the implications of the dimensions of leadership that matter into specific 'knowledge, skills and dispositions' (KSDs) of leaders (2009, Chapter 8). They identify four KSDs and present in helpful detail what each means, giving examples and non-examples, vignettes, mechanisms and supporting evidence. The four KSDs are:

- Ensure administrative decisions are informed by knowledge about effective pedagogy

- Analyse and solve complex problems

- Build relational trust

- Engage in open-to-learning conversations

Liebowitz and Porter (2019)

Liebowitz and Porter (2019) present a systematic review and meta-analysis of the effects of different principal behaviours on a range of outcomes, including student attainment. This review is methodologically stronger than many other reviews in the leadership field and adopts a more critical stance. In particular, it states clearly and unequivocally that the available evidence is purely correlational: "findings are based almost entirely on observational studies because the causal evidence base on school leadership behaviors is, to our knowledge, nonexistent"(2019, p. 5).

A further strength of Liebowitz and Porter's review is its conceptual clarity. They make a distinction between 'leadership behaviours' (descriptions of the things leaders actually do) and what are sometimes presented in the literature as 'leadership practices', such as "holding high expectations, achieving consensus among staff, or giving staff purpose for their work", which are really more of a description of the desired end point.

A common approach in reviews of leadership research is to report the associations between a range of different principal behaviours and student (or other) outcomes, and rank them in apparent order of importance.

Liebowitz and Porter do something more sophisticated and so provide a better answer to the question of whether it is appropriate to advise school leaders to prioritise one kind of behaviour (for example focusing on 'instructional management') over another. Two reasons why this is more complex than simply ranking the size of the correlations (or regression coefficients) are that these estimates are (a) themselves quite highly correlated within studies, and (b) estimated with quite low precision in most cases. They conclude that

> " ... prior literature may overstate the unique importance of instructional management as a tool to improve student achievement outcomes. At the least, we conclude that in our sample of studies the effects of time spent on, and skill in, instructional management on student achievement outcomes are similar in magnitude to time and skill in other types of principal behaviors." (Liebowitz & Porter, 2019, p. 26)

Although Liebowitz and Porter are far more robust than many other authors in this field in clarifying that causal inferences are unwarranted, even they appear to slip into accepting this interpretation. For example, in their conclusion they suggest that one way to leverage the relationship between the time principals spend on various leadership activities and student outcomes, in a context where principals already work more hours than they should, would be to hire additional school leaders, so enabling additional time to be spent on the behaviours that are associated with better outcomes. They present a calculation based on this estimated association as an 'upper bound', and list some key assumptions that underpin it, but even so, it seems like a rather implausible extrapolation from a thin evidence base.

Despite the fact that Liebowitz and Porter's review is methodologically stronger than others, there may be grounds for doubting its conclusions. For example, the headline findings of positive relationships between the five principal behaviours and student outcomes (standardised 'effect sizes' between 0.08 – 0.16[6]) could (and often would) just be reported as the point estimates and standard errors. Helpfully, however, Liebowitz and Porter provide a density plot that shows these estimates are based on a distribution of study effect sizes that has a positively skewed distribution (Liebowitz & Porter, 2019, p. 20, Table 3). The modal study 'effect' is close to zero, but the mean is brought up by a smaller number of studies with large positive 'effects'. It is hard to think of a reason why, if all available studies are to be thought of as randomly sampled estimates of a common effect, we would expect to see such a skew. An explanation based on reporting or publication bias seems much more likely.[7] If that is the case, then the headline figures appear much more questionable.

6   These correspond to estimated correlation coefficients of 0.04 – 0.08, using the formula $d=2r/sqrt(1-r2)$, cited in note 6, p.34
7   No funnel plot or other approach to estimating publication bias is provided, nor are studies coded for whether they were pre-registered.

Grissom et al. (2021)

The review by Grissom et al. (2021) for the Wallace Foundation has been mentioned above (p. 5) in relation to estimating the effects of different principals in the same school. A major part of this report consists of a systematic review and meta-analysis of studies that investigate relationships between characteristics or ratings of school principals and student outcomes. This review is clear about the limitations of the research base, acknowledging that:

> "few studies employed an experimental or quasi-experimental design with a credible strategy for making causal claims about the relationship between some aspect of leadership and a school outcome. Instead, most studies had correlational research designs, which have a weaker claim to causality (i.e., less internal validity) because of concerns about bias from omitted variables that may be driving the associations they observe." (Grissom et al., 2021, p. 82)

They did exclude studies with no credible control, but arguably set the bar lower than they might have done in a field where strong designs were more prevalent: "This minimal exclusion criterion eliminated a surprising number of studies." They also note that other "prominent meta-analyses of supposedly 'high-leverage' practices" have failed to exclude these weaker studies, "yet the language they employ often is causal, which can be misleading and imply—particularly to practitioners or those not trained in research methodology—that more has been demonstrated than actually has" (2021, p. 82). Additional methodological challenges the authors draw attention to concern generalising from the potentially unrepresentative contexts studied, and the almost total absence of any replication of findings: "we observed dizzying variation in what factors leadership studies considered, how those factors were operationalized, and the approaches the studies employed for analyses. Even among studies of the same topic, we seldom encountered two studies using the same measurement tools, or studies that replicated an earlier result" (2021, p. 55) Two further strengths of this review are the inclusion of case studies and other qualitative work and the triangulation of claims from the quantitative analysis against this evidence, and a detailed and unrelenting focus on equity and what leaders can do to promote it.

Although Grissom et al.'s review is about as methodologically strong as it could be, the limitations of the available studies mean that its conclusions should be treated with some caution. Typically, the studies with the strongest designs for causal claims (e.g., randomised controlled trials) are evaluations of well-defined interventions, such as a coaching programme or structured teacher observation programme.

The principal's exact role in these interventions is not always clear and many other details are likely to be crucial to their success, so it is a bit of a leap to give general advice to principals to introduce 'coaching' or 'teacher evaluation', for example. Nevertheless, they suggest that effective principles do four key things (Grissom et al., 2021, pp. 58–72):

Engaging in **instructionally focused interactions** with teachers:

- Teacher evaluation: collecting valid information; interpreting appropriately; getting buy-in
- Teacher feedback, coaching and other professional learning: giving actionable, constructive, robust feedback; supporting coaching; aligning professional learning opportunities with goals
- Using data to drive instructional improvement: monitoring student progress; managing performance

Building a productive **school climate**:

- Promoting behaviours such as collaboration, engagement with data and teacher (distributed) leadership
- Promoting values and beliefs among staff and community characterised by trust, press for continuous improvement, academic optimism (high expectations and efficacy)
- Promoting working conditions that foster wellbeing and safety

Facilitating productive **collaboration** and professional learning communities:

- Shared goals, resources, responsibilities and accountability
- Professional learning communities that are structured, supported and focused on instruction

Managing personnel and resources **strategically**:

- Optimisation and management of time and resources, aligned with goals
- Strategic teacher hiring, using the best available information
- Strategic assignment and placement of teachers
- Retaining high-performing teachers (and allowing lower performers to leave)

They also suggest that these practices are underpinned by three broad skills (Grissom et al., 2021, pp. 54–57).

Focus on **instruction**:

- Expertise about high-quality instruction to observe and evaluate teaching, identify effective practices
- Responsive and actionable feedback
- 'Orchestration' of professional learning, recognising and selecting what is high-impact

Interactions with **people**:

- Caring: offer support in safe and nurturing environments
- Communicate effectively: purposeful and open communications, able to have challenging conversations, building shared expectations (with teachers, parents, students)
- Trust: respect, empowerment/autonomy, reliability/competence

Manage the **organisation**:

- Operations: budgets, resources, facilities, staffing (and safety?)
- Data use: monitoring, quality assurance
- Strategic thinking: "set goals and think strategically about how to harness available resources to meet those goals"

Grissom et al. (2021) also list some ineffective leadership practices, not supported by evidence:

- Classroom walk-throughs (for 'informal' monitoring or just to appear visible)
- Post-observation checklists for giving feedback
- Licensure examinations for school leaders

The Dynamic Model of Creemers and Kyriakides

We presented quite a detailed summary of this model (see, e.g., Creemers & Kyriakides, 2011) in an appendix to the *Great Teaching Toolkit: Evidence Review* (Coe et al., 2020, pp. 54–60), because of its contribution to our knowledge of classroom practices that promote student learning. The model goes beyond classroom practice and also includes school-level factors, and higher levels. The Dynamic Model is distinctive for its strong (and broadly-based) theorisation, the rigour of its operationalisation in specific, high-quality measures, and the robustness of the empirical testing to which it has been subjected. At the school level, the Dynamic Model concentrates on school policy and actions that focus on two areas: teaching practices and the learning environment. For each of these, it measures both the things people do to promote them, and the approaches they use to evaluate them and respond to the results. At school level, the key measures are:

Teaching practices, for which the direct measures are:

- Quality of teaching
- Learning opportunities
- Quantity of teaching

Learning environment, for which the direct measures are:

- Collaboration between teachers
- Relations with parents
- Student behaviour outside the classroom
- Resources

(And for each of 'teaching practices' and 'learning environment', there is a measure of the extent to which the school evaluates and responds.)

In relation to the importance of school environment, the key study is probably Kyriakides and Creemers (2012). This reports the analysis of data from 50 primary schools in Cyprus, from which a wide range of data was collected, including attainment in mathematics and Greek language. The measures are robustly validated and analysed, using multilevel structural equation models. The analysis shows that the model that best fits the data includes both direct and indirect effects on student achievement. Achievement gains are predicted by all the seven direct measures listed above, but the school's performance on those measures is also predicted by the extent to which that feature is evaluated and responded to within the school.

In other words, there is evidence that internal formative evaluation of a school's teaching and environment is associated with better teaching and environment across the school, which in turn is associated with higher attainment, after adjusting for factors such as prior attainment, socioeconomic status, gender and individual classroom quality.

In the last decade or so, the focus of the researchers on the Dynamic Model has turned to developing and evaluating programmes of school improvement based on the model: the Dynamic Approach to School Improvement (DASI). The approach is strongly grounded in theory, recognises the need for individual schools to develop their own strategies, and to be flexible and responsive in doing so, but supports them with technical and research expertise to co-construct their strategy, and build in both formative and summative evaluation. Each school is assigned an Advisory and Research (A&R) Team, who work closely with them over typically one year. At least five randomised controlled trials, in a range of countries, have found substantial positive effects of DASI on attainment, equity and bullying reduction (Kyriakides et al., 2019).

Claim 5: Some school-level characteristics predict school and teacher growth

A smaller number of studies have longitudinal data that allows them to identify relationships between school-level characteristics over time and successive observations of student attainment, estimating school and teacher effects in subsequent years. Among these, two stand out. The first is Kraft and Papay's (2014) analysis of ten years of value-added attainment data, linked to individual teachers and matched at school level to staff surveys of school environment. The second is a strand of work by Bryk and colleagues (A. S. Bryk et al., 2010) over seven years in elementary schools in Chicago, again with detailed and repeated staff surveys linked to school attainment data.

Kraft and Papay (2014)

This study found substantial variation in the improvement trajectories of teachers, using value-added models with teacher fixed effects. The strongest predictor of this heterogeneity in growth rates was a measure of the school environment, derived from teacher surveys over three time points. Items in the 24-item survey were grouped under six sub-headings, but the final model used a weighted sum of all items as a single general measure of professional environment:

- Order and discipline: The extent to which the school is a safe environment where rules are consistently enforced and administrators assist teachers in their efforts to maintain an orderly classroom.

- Peer collaboration: The extent to which teachers are able to collaborate to refine their teaching practices and work together to solve problems in the school.

- Principal leadership: The extent to which school leaders support teachers and address their concerns about school issues.

- Professional development: The extent to which the school provides sufficient time and resources for professional development and uses them in ways that enhance teachers' instructional abilities.

- School culture: The extent to which the school environment is characterised by mutual trust, respect, openness, and commitment to student achievement.

- Teacher evaluation: The extent to which teacher evaluation provides meaningful feedback that helps teachers improve their instruction, and is conducted in an objective and consistent manner.

While the six components seem conceptually distinct, and perhaps could even be split further, Kraft and Papay found reasonable support for treating teachers' perceptions of professional environment as a unidimensional construct.[8]

Bryk et al. (2010)

Bryk et al. (2010) report an analysis of a major school reform project in Chicago in the 1990s. They collected student attainment data for students in 390 urban public elementary schools each year, together with survey responses from students and teachers at three time points (1991, 1994 and 1997). The surveys were designed around a conceptual framework and theory of change (outlined in A. S. Bryk et al., 2010, Chapter 2) about how the classroom interactions between teachers, students and subject matter depend on features such as student motivation and participation, learning time, and key resources and supports. These classroom features are in turn supported by the school's learning climate, its professional capacities, the nature of instructional guidance, and its ties with parents and community—with school leadership as the overarching driver of improvement.

The team developed the idea of 'essential supports': the underpinning requirements for a school to be able to improve. They identified five main essential supports—though of course, these are broad headings, subject to potentially arbitrary lumping/splitting choices, and over time, the exact headings have changed.

8   For example, they report internal consistency of the composite scale as above 0.9, and school-level intra-cluster correlation of around 0.3. The analysis used a weighted average of all 24 responses, using weights from the first principal component. In models where each component was entered separately, the coefficients of each were very similar (see Kraft & Papay, 2014, supplementary appendices).

In Bryk et al. the five, which reflect the 1997 version of the teacher and student surveys, are (from A. S. Bryk et al., 2010, p. 82, Table 3.1):

School leadership

- Inclusive leadership—degree to which teachers viewed their principal as an inclusive, facilitative leader, focused on parent and community involvement and creating a sense of community in the school.

- Instructional leadership—degree to which teachers viewed their principal as setting high standards and exercising leadership for instructional reform.

- Teacher influence—a measure of the extent of teachers' involvement in school decision making.

- LSC contribution—teachers' ratings of the effectiveness of the Local School Council.

- Program coherence—teachers' reports about the quality of implementation and coordination of programs within the school.

- SIP implementation—teachers' evaluations of the school improvement plan and its centrality to the school's efforts to improve student learning.

Parent-community ties

- Teacher outreach to parents—teachers' assessments of their efforts to develop common goals and understandings with parents and work together to strengthen student learning.

- Parent involvement in the school—teachers' reports about how regularly parents pick up report cards, attend parent-teacher conferences, attend school events, and other activities.

Professional capacity (work orientation)

- Teacher orientation toward innovation—teachers' assessments about whether their colleagues are continually learning, seeking new ideas, and have a can-do attitude.

- School commitment—teachers' reports of how loyal and committed they are to the school.

Student-centered learning climate (safety and order)

- Safety—students' perceptions of personal safety inside and outside the school and traveling to and from the school.

- Classroom disruptions—teachers' reports of disruptions due to students' behavior and administrative interruptions.

Instructional guidance (curriculum alignment)

- Assesses the pace with which new math content is introduced into the school's curriculum across the elementary grades and how well this aligns with established grade-level skills and knowledge.

By 2020, the 5Essentials survey instruments are being offered to schools serving K-12, beyond just Chicago, and collecting a wider range of student outcomes (Hart et al., 2020).[9] The names of the five essential supports, and their component indicators, have changed slightly:[10]

- Effective leaders (program coherence T; teacher-principal trust T; teacher influence T; instructional leadership T)

- Collaborative teachers (collaborative practices T; collective responsibility T; quality professional development T; school commitment T; teacher-teacher trust T)

- Involved families (teacher-parent trust T; parent involvement in school T; parent influence on decision-making in schools T)

- Supportive environment (peer support for academic work K-8, S; academic personalism K-8, S; safety S; student-teacher trust S; school-wide future orientation HS, S; expectations for post-secondary education HS, T)

- Ambitious instruction (English instruction S; math instruction S; academic press T; quality of student discussion T)

Bryk et al. (2010) use a sophisticated approach to identifying school-level trends of improvement or decline in three outcomes: reading, mathematics and attendance. For reading and mathematics, they place scores on the different versions of the Iowa Test of Basic Skills (ITBS) on a common scale using Rasch analysis, so are able to estimate the amount that each child has learnt each year. At the school level, these trajectories are adjusted to take account of the school's context and starting point, so that an 'improving' school means one where the improvement over time in the average rate of learning is greater than in similar schools. Trends in attendance are also adjusted for student composition, to allow like-with-like comparisons. Given that the overall trend during this period was improving, they categorise schools in the top quartile of rising performance on each of the three indicators as 'improving' and those in the bottom quartile as 'stagnant'.

Bryk et al. present substantial evidence that a school's strength on each of the Five Essential Supports is a good predictor of subsequent improvement or stagnation in all three outcomes in future years.

---

9   More details about the survey and professional support, materials, etc., can be found at **https://uchicagoimpact.org/our-offerings/5essentials**

10 These are the Five Essential Supports, together with the 20 measures that underpin them. S denotes student survey; T denotes teacher survey; K-8 is only for grades K-8; HS is High School only.

Within this general pattern there are some differences: for example, teacher reports about safety and order were the strongest single predictor of improvement in attendance, but the weakest predictor for reading and mathematics. Each of the other four supports was a very strong predictor of improvement in mathematics and a strong predictor of improvement in reading. Overall, trends in attendance were less strongly predicted by the survey results than trends in the attainment measures. There were also some non-linear relationships, where a low rating (bottom quartile) on one of the survey scales made improvement on an outcome very much less likely, but a high score (top quartile) made it only somewhat more likely, but these are much smaller effects.

The justification for describing these support factors as 'essential' is that, if we allow for the limitations of measurement, the strength of the relationships is such that pretty much no schools that were weak in any of the essential supports showed strong improvement in any of the outcomes. The small number of schools actually observed to break this rule is compatible with what we would expect, given plausible assumptions about indicator unreliability, construct underrepresentation and temporal instability (A. S. Bryk et al., 2010, p. 92). Moreover, Bryk et al. argue that the dependencies among the five essential supports are such that an ongoing weakness in any one of them will undermine the strength and impact of the others. While many schools will be in a state of flux, and strategically may choose to focus on a subset at any given time, the combination of strength in all the five essential supports must be the goal if they want to achieve sustainable improvement in student outcomes.

The claim that leadership strength is the driver of change in all the other supports is tested by comparing schools that were matched at baseline on their ratings on the other four essential supports but differed in their leadership rating, to see how later growth in those indicators varied. A pattern of strong leadership predicting subsequent improvement in the other indicators was seen over both time intervals (between surveys in 1991-4 and 1994-7); schools with weak leadership ratings were much less likely to improve in their other supporting factors.

Claim 6: There is still a lot we don't know about school leadership and environment

The previous five claims have tried to summarise what we do know (albeit with some significant caveats) about school leadership and environment, based on what we have judged to be among the best available studies currently. Perhaps partly to reinforce and clarify those caveats, this section highlights areas that are more negative: things we do not know.

The malleable, well-defined behaviours of school headteachers that make a difference

The evidence summarised above does give us some pointers to school and leader characteristics that are associated with higher student attainment and the kinds of interventions that leaders may be able to mobilise to achieve a positive impact. However, there seems to be nothing in any of this literature that: (a) identifies specific, well-defined behaviours or actions school leaders could take; (b) demonstrates that leaders can deliberately choose or learn to do these things if they weren't already; and (c) shows that doing so results in improvements in valued outcomes, such as students' overall attainment or equity. Without this kind of evidence, there is no direct path to improving the practices of leadership.

The size and nature of the contribution of school leaders who are not headteachers (e.g., assistant headteachers)

Although it is common in leadership discourse to view school leadership as a shared endeavour and endorse the importance of senior leadership teams and middle leaders, none of the evidence we have summarised seems to say much about the scale of impact of their contribution, or the characteristics and behaviours of senior and middle leaders that make a difference. A review by Goldring et al. (2021) shows that although some studies have considered the role of assistant principals, there is little evidence about how they influence key outcomes. Of course, that is not to say that no claims are ever made about what these leaders should do or how important they are—research and discourses on leadership are full of claims that are meaningless or wrong. But it is not clear that there is any robust empirical support for any of these claims.

How to train and support headteachers and school leaders to be more effective

The systematic review by Grissom et al., mentioned above, also summarises the evidence from evaluations of training, induction or in-role support programmes for school principals (2021, pp. 79–80). Given the worldwide investment in training school leaders—and its importance—it is something of an indictment of the research in this area to see that their systematic search found only ten evaluations of eight interventions.[11]

Grissom et al.'s interpretation of this evidence is that interventions to support early-career or new principals generally show positive effects, while those for mixed groups or more experienced principals do not. However, the only two randomised designs are both in the latter group (both show no effect), and the evaluations in the former group do not satisfactorily rule out selection effects: principals who went through the programme were more effective, but we cannot say how much of that was due to the programme, and how much to pre-existing differences.[12]

Hence, a more critical interpretation would be that there is no convincing evidence that any training programme for school leaders has led to any discernible benefit for students, and that we do not really have any secure knowledge about how to enhance the effectiveness of school leaders. Liebowitz and Porter conclude that "the contribution of principal training programs to principals' influence on student outcomes appears to be modest at best and presents measurement challenges" (2019, p. 3).

Anything about the size and distribution of teacher or headteacher effects from different contexts (e.g., UK)

As ever, available research comes from a limited range of contexts, from which it may be problematic to extrapolate if the question we want to answer is, "Does this finding apply in my school?" For example, the majority of the evidence we have considered is from studies conducted in North America; notable exceptions are the Dynamic Model, which has been developed and evaluated in Europe, and some sections of Robinson et al.'s (2009) review that specifically included studies from New Zealand.

Most of the quantitative studies with student attainment data are conducted in primary (elementary) schools. There are a number of reasons we might expect leadership to operate differently in secondary (high) schools: the size of the organisation; differences in curriculum, including subject specialism; differences in the age and development of the students; differences in typical parental engagement in the school; etc.

---

11   Grissom et al.'s (2021) systematic review was limited to studies in the US since 2000
12   See Methodological challenges in school leadership research, p18, for more details of this argument.

Perhaps even more limited is the range of attainment data in all these studies, which seems to be exclusively from students in grades 3 to 8 in either mathematics or English (or other home) language. Application to younger or older students, or to other curriculum areas, currently rests on the assumption that they will be the same.

The extent to which teacher or principal effects depend on interactions with key factors

Among a list of questions to which we do not yet have good answers are:

How do estimates of headteacher or teacher effects on attainment depend on the characteristics of the particular classes or students they teach? For example:

- Will the same headteacher or teacher generally look worse in a more challenging school?

As well as the direct relationships between school-level characteristics or headteacher behaviours and attainment, are there interactions among different predictors? For example:

- Does a safe and ordered environment matter more in high-poverty schools?

- How do perceptions of trust and accountability interact to influence attainment? (Do you need both to be strong, or is there an independent effect of each without the other?)

How do estimates of teacher and headteacher effects relate to effectiveness estimates of previous and subsequent teachers and leaders (on the same students, in the same subjects), or to estimates for peers concurrently teaching different students or subjects?

References

Bartanen, B. (2020). Principal quality and student attendance. *Educational Researcher, 49*(2), 101–113. https://doi.org/10.3102/0013189X19898702

Bartanen, B., Husain, A. N., & Liebowitz, D. D. (2022). *Rethinking principal effects on student outcomes.* https://doi.org/10.26300/r5sf-3918

Branch, G. F., Hanushek, E. A., & Rivkin, S. (2013). School leaders matter: Measuring the impact of effective principals. *Education Next, 13*(1), 62–69. https://www.educationnext.org/school-leaders-matter/

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2009). *Estimating principal effectiveness* (No. 32). https://files.eric.ed.gov/fulltext/ED509685.pdf

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). https://doi.org/10.3386/w17803

Bryk, A., Harding, H., & Greenberg, S. (2012). Contextual influences on inquiries into effective teaching and their implications for improving student learning. *Harvard Educational Review, 82*(1), 83–106. https://doi.org/10.17763/haer.82.1.k58q7660444q1210

Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan, 91*(7), 23–30. https://doi.org/10.1177/003172171009100705

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organising schools for improvement: Lessons from Chicago.* The University of Chicago Press.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review, 104*(9), 2633–2679. https://doi.org/10.1257/AER.104.9.2633

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy, 11*(3), 283–309. https://doi.org/10.1162/EDFP_a_00184

Coe, R., Rauch, C. J., Kime, S., & Singleton, D. (2020). *Great teaching toolkit: Evidence Review.*

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review, 31*(1), 92–109. https://doi.org/10.1016/j.econedurev.2011.09.001

Creemers, B. P. M., & Kyriakides, L. (2011). *Improving quality in education: Dynamic approaches to school improvement.* Routledge. https://doi.org/10.4324/9780203817537

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics/Revue Canadienne d'économique, 47*(2), 634–663. https://doi.org/10.1111/caje.12086

Dhuey, E., & Smith, J. (2018). How school principals influence student learning. *Empirical Economics, 54*(2), 851–882. https://doi.org/10.1007/S00181-017-1259-9/TABLES/10

Ericsson, K. A. (2009). *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (K. A. Ericsson, Ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511609817

Goldring, E., Rubin, M., & Herrmann, M. (2021). *The role of assistant principals: Evidence and insights for advancing school leadership.* https://www.wallacefoundation.org/knowledge-center/pages/the-role-of-assistant-principals-evidence-insights-for-advancing-school-leadership.aspx

Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research.* http://www.wallacefoundation.org/principalsynthesis.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis, 37*(1), 3–28. https://doi.org/10.3102/0162373714523831

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources, XXXIX*(2), 326–354. https://doi.org/10.3368/jhr.XXXIX.2.326

Hart, H., Young, C., Chen, A., Zou, A., & Allensworth, E. M. (2020). Supporting school improvement: Early findings from a reexamination of the "5Essentials" survey. Research Report. In *University of Chicago Consortium on School Research.* University of Chicago Consortium on School Research. 1313 East 60th Street, Chicago, IL 60637. Tel: 773-702-3364; Fax: 773-702-2010; Web site: http://consortium.uchicago.edu/. https://eric.ed.gov/?id=ED608120

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics, 1*(4), 85–108. https://doi.org/10.1257/APP.1.4.85

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record: The Voice of Scholarship in Education, 114*(10), 1–39. https://doi.org/10.1177/016146811211401004

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement. *American Educational Research Journal, 53*(5), 1411–1449. https://doi.org/10.3102/0002831216667478

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis, 36*(4), 476–500. https://doi.org/10.3102/0162373713519496

Kyriakides, L., Charalambous, E., Creemers, B. P. M., Antoniou, P., Devine, D., Papastylianou, D., & Fahie, D. (2019). Using the dynamic approach to school improvement to promote quality and equity in education: A European study. *Educational Assessment, Evaluation and Accountability, 31*(1), 121–149. https://doi.org/10.1007/S11092-018-9289-1/TABLES/9

Kyriakides, L., & Creemers, B. P. M. (2012). School policy on teaching and school learning environment: Direct and indirect effects upon student outcome measures. *Educational Research and Evaluation, 18*(5), 403–424. https://doi.org/10.1080/13803611.2012.689716

Laing, D., Rivkin, S., Schiman, J., & Ward, J. (2016). *Decentralized governance and the quality of school leadership* (No. 22061). https://doi.org/10.3386/w22061

Liebowitz, D. D., & Porter, L. (2019). The effect of principal behaviors on student, teacher, and school outcomes: A systematic review and meta-analysis of the empirical literature. *Review of Educational Research, 89*(5), 785–827. https://doi.org/10.3102/0034654319866133

Papay, J. P., & Kraft, M. A. (2017). Developing workplaces where teachers stay, improve, and succeed. In E. Quintero (Ed.), *Teaching in Context: The Social Side of Education Reform.* Harvard University Press.

Robinson, V., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why.* https://www.educationcounts.govt.nz/__data/assets/pdf_file/0015/60180/BES-Leadership-Web-updated-foreword-2015.pdf

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Seashore Louis, K., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Investigating the links to improved student learning: Final report of research findings.* https://conservancy.umn.edu/bitstream/handle/11299/140885/Learning-from-Leadership_Final-Research-Report_July-2010.pdf?sequence=1&isAllowed=y