Australian Education Research Organisation

AISWA

Sustaining professional learning outcomes: Evaluating the effectiveness of 'booster' interventions

Findings from a randomised trial

A collaboration between AERO and the Association of Independent Schools of Western Australia

October 2025

The Australian Education Research Organisation (AERO) is Australia's national education evidence body, working to achieve excellence and equity in educational outcomes for all children and young people.

Acknowledgement

AERO's work is made possible by the joint funding it receives from Commonwealth, state and territory governments.

AERO is grateful to the teachers, principals, and schools that participated in this study. AERO would also like to acknowledge Kris Stafford and Sarah Wells from the Association of Independent Schools of Western Australia (AISWA) for their partnership, delivery of professional learning, feedback and contributions to this study.

AERO would also like to thank the 2 academic reviewers who provided expert reviews of this content.

Acknowledgement of Country

AERO acknowledges the Traditional Custodians of the lands, waterways, skies, islands and sea Country across Australia. We pay our deepest respects to First Nations cultures and Elders past and present. We endeavour to continually value and learn from First Nations knowledges and educational practices.

Authors

Oscar Kovacs, Emily Greaves, Ioana Ramia, Julie People, Deborah Bradford and Wendy Moore

How to cite

Kovacs, O., Greaves, E., Ramia, I., People, J., Bradford, D. & Moore, W. (2025). *Sustaining professional learning outcomes: Evaluating the effectiveness of 'booster' interventions: Findings from a randomised trial.* Australian Education Research Organisation. https://www.edresearch.edu.au/research/research-reports/sustaining-professional-learning-outcomes-evaluating-effectiveness-booster-interventions

Publication details

ISBN 978-1-923066-65-6 (online)

Cover image: AERO

Contents

Figures

Tables

Glossary

Term	Definition
Assessment and Evaluation Research Centre (AERC)	AERC is part of the University of Melbourne, conducting research focused on educational measurement and assessment.
Australian Education Research Organisation (AERO)	AERO is Australia's independent education evidence body. AERO is a ministerial-owned company governed by a board. AERO conducts research and shares knowledge to promote better educational outcomes for Australian children and young people.
Association of Independent Schools of Western Australia (AISWA)	AISWA is the peak body for Western Australian independent schools. Providing professional learning services and programs is one of AISWA's core functions. These services are designed to support educators in enhancing their skills and knowledge, supporting the design and delivery of high-quality education for students.
Education Endowment Foundation (EEF)	The EEF is a United Kingdom-based charity focused on breaking the link between family income and educational achievement. It aims to support education settings to improve teaching and learning through better use of evidence.
Evidence-based practice (EBP)	Evidence-based practice(s) are pedagogical approaches backed up by research evidence. This means there is broad consensus from rigorously conducted evaluations that they work.
Evidence-based strategies and techniques	Teaching techniques and strategies are the components that make up evidence-based practice(s). For instance, removing scaffolding based on evidence of student learning is a teaching technique within the broader evidence-based practices of explicit instruction and formative assessment.
High Impact Practices program (HIPs program)	The HIPs program is AISWA's evidence-informed professional learning program, focused on teacher use of evidence-based practices.

Summary

Professional learning is a commonly used strategy to help teachers develop the skills and knowledge they need to use evidence-based practices in the classroom. Recent research has established 14 building blocks, or mechanisms, that can contribute to the efficacy of professional learning (Sims et al., 2021). These include revisiting prior learning, practical social support, managing cognitive load, goal setting and modelling teaching practices. Professional learning interventions that contain more of these mechanisms tend to be more effective, especially when a balance of varied mechanisms is incorporated. Despite strong evidence for the importance of these mechanisms, there are still gaps in our knowledge about how to apply them to make professional learning as effective as possible and achieve lasting change. There are also gaps in our knowledge about how to apply the mechanisms at scale so that a greater number of teachers and students can benefit.

Prior research has shown that brief or light-touch professional learning interventions are typically ineffective in changing teacher practice or student outcomes. Approaches that are more effective tend to be sustained in duration or have a follow-up component such as coaching, engaging with communities of practice or receiving formal external support. The drawback to these approaches is that they tend to be resource-intensive, limiting their accessibility and scalability.

Refresher interventions or 'boosters' that are delivered as a follow-up to other professional learning programs may be an effective and scalable way to improve professional learning and create lasting change. There is emerging evidence that boosters can improve long-term outcomes, and they are particularly well-placed to incorporate mechanisms like revisiting prior learning and practical social support. Unlike other light-touch interventions, boosters are not delivered in isolation, and as a follow-up to other professional learning, they may be an effective way to leverage these mechanisms at scale. Rigorous causal research is consequently needed to determine whether scalable boosters can improve professional learning and create sustained improvements in teacher practice and student outcomes.

To examine the effectiveness of scalable booster interventions as a follow-up to professional learning, we conducted a 2-arm randomised trial in partnership with the Association of Independent Schools Western Australia (AISWA). The study was conducted in 17 primary and secondary schools in Western Australia. It compared teachers who received a scalable booster intervention after completing a professional learning program with those who received the initial professional learning only. A total of 107 teachers and 17 school leaders consented to participate in the study.

The professional learning program completed by all participating teachers was AISWA's High Impact Practices (HIPs) program, an evidence-informed program incorporating effective mechanisms of professional learning, which focuses on instructional strategies that enhance teaching practices and student outcomes. Following the HIPs program, schools were randomly allocated to either receive the booster intervention approximately 6 months later or to continue with business as usual as a comparison group.

The booster intervention was co-designed by AISWA and the Australian Education Research Organisation (AERO) as a scalable follow-up to the HIPs program. It was based on the available evidence of effective booster interventions and the effective mechanisms of professional learning. To support lasting improvements, the booster was primarily designed around revisiting prior learning, as this mechanism has been strongly associated with long-term outcomes. However, in line with the research, we incorporated a balance of other effective mechanisms of professional learning to maximise the potential of the booster.

We collected data on teachers' reported confidence in using evidence-based practices, their reported quality of use of evidence-based practices and perceived student outcomes. Data was collected at the beginning of the study as a baseline, immediately after the initial professional learning program, and 6 months after the booster intervention (one year after the initial professional learning).

Key findings

- **The scalable booster intervention tested in this study was not effective.** While teacher outcomes improved over the course of the study, this improvement was common to both study groups, with no additional benefit for teachers who participated in the booster. Incorporating effective mechanisms of professional learning did not make this light-touch intervention effective.
- **Peer support and check-ins were valuable components of the booster that could support professional learning.** Despite the lack of measurable improvement, teachers were highly positive about their experience of the booster, reporting that several components were helpful in improving their practice. The most helpful of these included responsive, collaborative support and formalised check-ins following professional learning. These were reported to help troubleshoot the application of learnings to practice, refresh content knowledge and sustain momentum. Although these components did not result in measurable improvements within the constraints of a scalable booster, they are strongly aligned with the mechanisms of effective professional learning and could be integrated into existing school structures to support outcomes following professional learning.

Overall, current evidence from this and other studies shows that light-touch, scalable professional learning interventions do not meaningfully enhance teaching practice, either on their own or after a period of professional learning. Sustaining change takes time and resources, and so far, there are no demonstrated shortcuts to improving outcomes and achieving lasting change. Based on our results, it appears that light-touch interventions do not fail because of a lack of effective mechanisms, but because keeping them scalable makes them too constrained to leverage these mechanisms effectively. It is possible that future research could identify a more effective form of scalable booster intervention, but unless this evidence emerges, schools and systems should avoid light-touch interventions and focus on leveraging the effective mechanisms of professional learning through quality interventions that are sustained in duration.

1. Background

This report presents the findings of a study conducted by the Australian Education Research Organisation (AERO) in partnership with the Association of Independent Schools of Western Australia (AISWA). The study focuses on how teachers can implement and maintain practice changes following professional learning. Research suggests that effective professional learning interventions rely on specific mechanisms and tend to be more successful when they are sustained in duration. However, it can be difficult to deliver this type of professional learning at scale. There is preliminary evidence that brief refresher training after an initial block of professional learning can support long-term outcomes for teachers and students, but evidence from Australian education settings is limited. This project tested whether providing a light-touch, scalable professional learning 'booster' to revisit prior learning after the completion of an initial professional learning program can improve long-term teacher outcomes.

1.1. Rationale

Professional learning plays an important role in supporting teachers to develop the skills and knowledge needed to use evidence-based practices. Recent research has established many of the building blocks, or mechanisms, that can make professional learning effective (Sims et al., 2021). However, there are gaps in knowledge about how best to apply these mechanisms, how to support teachers to sustain improvements in practice over time and how to achieve this at scale (Moore et al., 2024; Strunk et al., 2016). Evidence is needed to fill these gaps so we can provide actionable guidance on how to design and deliver professional learning programs that are effective, scalable and create lasting change.

1.2. Effective professional learning relies on specific mechanisms

A recent systematic review and meta-analysis by the Education Endowment Foundation (EEF) showed that effective professional learning interventions for teachers typically include a range of different mechanisms (Sims et al., 2021). The authors of this study identified 14 mechanisms of effective professional learning, which they grouped into 4 categories:

1. instil insight (manage cognitive load, revisit prior learning)
2. motivate goals (goal setting, credible source, praise/reinforcement)
3. teach techniques (instruction, practical social support, modelling, feedback, rehearsal)
4. embed practice (prompts/cues, action planning, self-monitoring, context-specific repetition).

To be included, each mechanism needed to have a causal explanation for how it contributes to professional learning, and it had to be domain-general, drawn from research in multiple contexts in addition to teacher professional learning.

The authors compared the efficacy of 205 professional learning interventions for teachers, accounting for inclusion of these mechanisms. They found that professional learning interventions incorporating more of these mechanisms tended to be more effective, but also, critically, that interventions with mechanisms from each of the 4 categories (termed 'balanced designs') were particularly impactful.

While there is compelling evidence for the existence of these mechanisms and their contribution to effective professional learning, the prevalence of short-term research within the evidence base from which they were derived means that they may not generalise as well to the longer term. Most professional learning research is skewed towards relatively short-term outcomes such as those collected during or immediately after program delivery, as these are easier and cheaper to evaluate in empirical work, and obtaining some evidence of efficacy can be a precondition to program success and engagement (Fullan, 2006; Zehetmeier & Krainer, 2011). As a result, there is a lack of evidence around how to create lasting improvements through professional learning, even with our improved knowledge of the underlying mechanisms.

1.3. Sustaining improvements from professional learning is important, but challenging

Professional learning that exposes teachers to new practices is generally not enough to lead to lasting change. Teachers need support to embed new practices and maintain them over time, but effective approaches can be resource-intensive.

Following professional learning, momentum tends to wane, both over time and as competing priorities emerge, with teachers increasingly reverting to old habits unless a specific effort is made to sustain the new approach (Darling-Hammond et al., 2017; Moore et al., 2024). An additional challenge is that without longer-term follow-up or reinforcement, people tend to forget information and cease trying to adapt new learning to their context. This can happen quickly, especially when that information is complex, challenging to implement or poorly aligned with existing beliefs or practice (Kennedy, 2016; Riley-Tillman & Eckert, 2001; Rose & Church, 1998; Thoonen et al., 2011; Wei et al., 2010). In these cases, even high-quality professional learning may have only a short-term effect, and practices and knowledge gained from professional learning programs can fail to generalise.

Approaches that have shown promise in creating lasting improvements include professional learning that is sustained in duration or has a follow-up component, such as coaching, engaging with communities of practice or receiving formal external support. The drawback to these approaches is that they tend to be resource-intensive, making them a weaker return on investment and limiting their accessibility and scalability (Edwards Groves & Rönnerman, 2013; Elias, 2010; Fixsen et al., 2005; Harris & Sass, 2011; Hill et al., 2013; Kraft et al., 2018; Prenger et al., 2017).

1.4. Scalable and accessible light-touch interventions are appealing, but generally ineffective on their own

One-off, easy to deliver professional learning interventions are appealing and highly scalable. Meta-analytic approaches have indicated that most professional learning takes this form, but also that these sorts of 'light-touch' approaches tend to produce little improvement in teaching practice (Darling-Hammond et al., 2017; Garet et al., 2001; Wei et al., 2010; Weiss & Pasley, 2006). The EEF tested a range of light-touch interventions to professional learning over a series of research trials collectively known as the 'Literacy Octopus' trials (Lord et al., 2019a, 2019b). These interventions, which aimed to improve teaching practice, ranged from providing resources that included guidance on translation into practice, to resources supplemented by events, webinars, workshops or peer observations.

Engagement with the interventions was good, and the majority of participating teachers reported that they felt that it had improved their practice. However, none of these professional learning interventions had any meaningful effect on teacher or student outcomes, either in the year of the trial or in the following year (Lord et al., 2019a, 2019b).

Light-touch interventions are easy to deliver at scale, but many of the 14 effective mechanisms of professional learning described by Sims et al. (2021) are inherently ill-suited to such brief or one-off interventions. For example, scalable, light-touch interventions are unlikely to contain opportunities to receive feedback, engage in action-planning and monitoring or in rehearsal and repetition. Similarly, a 2017 review by Darling-Hammond et al. identified 7 features that effective professional learning interventions tend to share. These included active learning, coaching, feedback and reflection, and sustained duration – all features that are also difficult to implement in brief interventions, and that scalable light-touch interventions will often lack. It is possible that scalable, light-touch interventions are ineffective because when they are used in isolation, they do not sufficiently incorporate the mechanisms of effective professional learning.

1.5. Professional learning could be supplemented by scalable, light-touch follow-up interventions to sustain practice change

Although ineffective on their own, scalable light-touch interventions may be useful as a 'booster' in sustaining instructional changes achieved through professional learning that utilises effective mechanisms. Boosters have a natural advantage in falling after initial professional learning programs, making it possible to have a formalised check-in at a point when forgetting content, running into translation issues or maintaining momentum are likely to be challenges. Following a professional learning program, booster interventions can provide an opportunity to revisit content, encourage reflection and collaborate with peers, helping to deepen learning and navigate the implementation of new practices (Hill & Papay, 2022). As short, externally facilitated sessions, they require less time, staffing, coordination and resources than a new sustained professional learning program or coaching program would entail. If they can produce meaningful improvements at scale, they may be a valuable component to build into professional learning initiatives.

A range of studies, while different in design, content and context, have demonstrated the potential of revisiting information through booster training to support and sustain improvements over time:

- Two small-scale studies in schools indicated that providing teachers with initial training and follow-up contact or refresher training had a positive effect on student outcomes when compared to initial training only (Anshu, 2019; Owen et al., 2021).
- Watson (2006) found an increase in teacher self-efficacy with the addition of online booster sessions after an intensive professional learning course.
- Dewing et al. (2014) found an improvement in skills following a booster and supervision in a different professional field (counselling).
- Frank and Kluge (2018) demonstrated that refresher interventions assist people of varying cognitive ability with retaining complex skills.
- Stanton (2015) demonstrated that a booster session delivered with high fidelity contributed to embedding and sustaining practice change in a sample of primary and lower secondary school teachers.
- Miller et al. (2014) found that individualised booster training was effective in re-establishing teacher performance at or above the level achieved after initial training in classroom management.

Despite the promising findings, more work is needed to determine whether there is a form of booster intervention that can be used to improve long-term outcomes without compromising its ability to be delivered at scale. The strength of the evidence from these studies varies, and not all the studies were conducted in educational or Australian contexts. In addition, the boosters themselves were highly diverse interventions and some relied on components that are resource-intensive, limiting their ability to be delivered at scale (e.g., individualisation, multiple sessions, broader staff involvement – see the Elements of successful boosters for a more complete list). While most of these boosters incorporate some effective mechanisms of professional learning, such as revisiting prior learning, they were not based on this new information and are unlikely to qualify as balanced designs. Designing balanced booster interventions with the effective professional learning mechanisms in mind has the potential to improve their efficacy.

We developed this study to test whether a scalable booster that incorporates as many effective mechanisms as possible can improve outcomes following an initial professional learning program.

Elements of successful boosters

The following are some of the elements of successful boosters we identified in the literature:

- **alignment** with the content covered in the initial professional learning (Stanton, 2015)
- **appropriate content focus** and delivery for participants' experience and roles (Chipchase et al., 2016)
- **individualised content** based on participant needs, as identified through evaluation of the initial professional learning sessions, field observations, or a pre-booster assessment (Anshu, 2019; Miller et al., 2014)
- **informal support and a shared context** from having multiple teaching staff from the same school attend together (Crowther & Cannon, 2002; Rusby et al., 2013)
- **interactive formats** are more successful than boosters based on traditional didactic approaches (Chipchase et al., 2016)
- **components prepared by participants** increase engagement and accountability (Crowther & Cannon, 2002)
- **involving support staff and school leadership** teams in the professional learning process and booster component (Anshu, 2019)
- **reinforcing basic skills** in booster sessions and encouraging participant research and unit planning (Watson, 2006)
- **multiple booster sessions**, as a single session follow-up is insufficient to significantly influence change (Chipchase et al., 2016).

2. The research study

This section provides a brief overview of the study design, sample and data sources. The primary outcome measures are then introduced, and program fidelity is discussed. See Appendix A for additional details of the research methodology.

2.1. Study overview

2.1.1. Study aims

The aims of this project were to:

1. examine the impact of receiving a professional learning booster on teachers' use of evidence-based teaching practices over time
2. explore participants' experiences of the booster and perceptions of which elements of the booster were most effective at facilitating practice change
3. identify barriers and enablers to changing practice after participation in the professional learning booster.

2.1.2. Study design

To examine the effectiveness of scalable booster interventions as a follow-up to professional learning, we conducted a randomised trial in partnership with AISWA. For this project, AISWA provided the professional learning and AERO conducted the evaluation. The study was a 2-arm cluster randomised trial, focused on teacher-level outcomes following a professional learning program. The study was conducted in primary and secondary schools in Western Australia, and compared teachers who received a booster intervention after professional learning with those who received the professional learning only, to isolate the relative impact of a booster as a follow-up.[1] A convergent mixed-methods design was used, in which the qualitative and quantitative data from each timepoint were analysed separately, then integrated to provide a more complete picture (Creswell & Clark, 2017).

The initial professional learning program was AISWA's High Impact Practices (HIPs) program, an evidence-informed program focusing on instructional strategies that enhance teaching practices and student outcomes (see the HIPs program overview and Appendix A.3.1 for details). The HIPs program incorporates many effective mechanisms of professional learning, and would be considered a 'balanced' design, as it includes mechanisms from each of the 4 categories. This program was provided to all participating teachers in both arms of the study from May to August 2023.

1 This study does not have a control group that participates in the study but does not receive any 'treatment'. We were primarily interested in the relative impact of a professional learning booster intervention, not a professional learning program that also included a booster intervention. The most appropriate group to use as a comparison was one where professional learning was provided without a booster.

HIPs program overview

HIPs is an evidence-informed program that focuses on teacher use of evidence-based practices. The HIPs program incorporates delivery methods known to be effective (e.g., coaching and over 20 hours of instruction time) as well as many of the effective mechanisms of professional learning identified by the EEF, which would qualify it as a balanced design (see Appendix A.3.1 for a breakdown of mechanisms).

The program is built around 4 core principles:

1. Effective teaching is informed by contemporary educational neuroscience research.
2. Effective teaching has been identified as the biggest determinant of student improvement in schools.
3. The practices outlined are relevant to all teachers in all stages of their professional practice.
4. Collegial learning supports the development of sustained and meaningful improvements in practice and student outcomes.

The HIPs program includes:

- required pre-course readings and online learning modules
- three face-to-face workshops covering the high-impact strategies
- post-workshop activities completed individually and as a group
- classroom observations with coaching and feedback by an AISWA HIPs facilitator (one or 2 sessions, either video recorded or live).

After completing the HIPs program, schools were randomly allocated to one of the two study arms: either receiving the booster intervention 6 months after completing the HIPs program or continuing with business as usual to act as a comparison group. We expected that business as usual following professional learning could include activities that have the potential to change teaching practices, but these were not the focus of our investigation and were controlled for by randomisation and comparison across groups.

The booster intervention was developed as a collaboration between AISWA and AERO. It was designed as a scalable follow-up to the HIPs program, based on available evidence for effective boosters and the effective mechanisms of professional learning (Sims et al., 2021 – see the Elements of successful boosters, the Booster intervention overview and Appendix A.3.2). The booster was built around revisiting prior learning, one of the effective mechanisms of professional learning that has been most strongly linked to improved long-term outcomes (Adesope et al., 2017; Moreira et al., 2019; Pastötter & Bäuml, 2014; Sims et al., 2021), but we ensured that it incorporated as many other mechanisms as possible to qualify as a balanced design and to support teachers to continue using evidence-based teaching practices over time (see Appendix A.3.2 for a breakdown of mechanisms). The booster intervention was delivered to teachers in the booster arm of the study from January to March 2024.

Booster intervention overview

The booster was a novel intervention, collaboratively designed by AERO and AISWA based on a review of relevant literature on refresher interventions (see the Elements of successful boosters). It was designed as a follow-up to the HIPs program, with directly aligned content and a selection of effective mechanisms that would qualify it as balanced professional learning (i.e., mechanisms from each of the instil insight, motivate goals, teach techniques and embed practice categories – Sims et al., 2021). It was structured around the mechanisms of revisiting prior learning, and incorporated as many other mechanisms of effective professional learning as possible while retaining the ability to be delivered at scale (see Appendix A.3.2 for details).

The booster intervention comprised two 90-minute online professional learning workshops delivered by an AISWA facilitator, approximately 6 months after the conclusion of the HIPs program. The 2 workshop sessions were delivered to each school individually, with a 2-week gap in between sessions. Teachers in the Booster group were tasked with completing unfacilitated peer support activities related to their learning after each workshop (e.g., lesson planning, mentoring or classroom observations) to sustain peer support and collaboration, which are core elements of the HIPs program and mechanisms of effective professional learning (revisiting prior learning and practical social support).

The booster sessions provided an opportunity to:

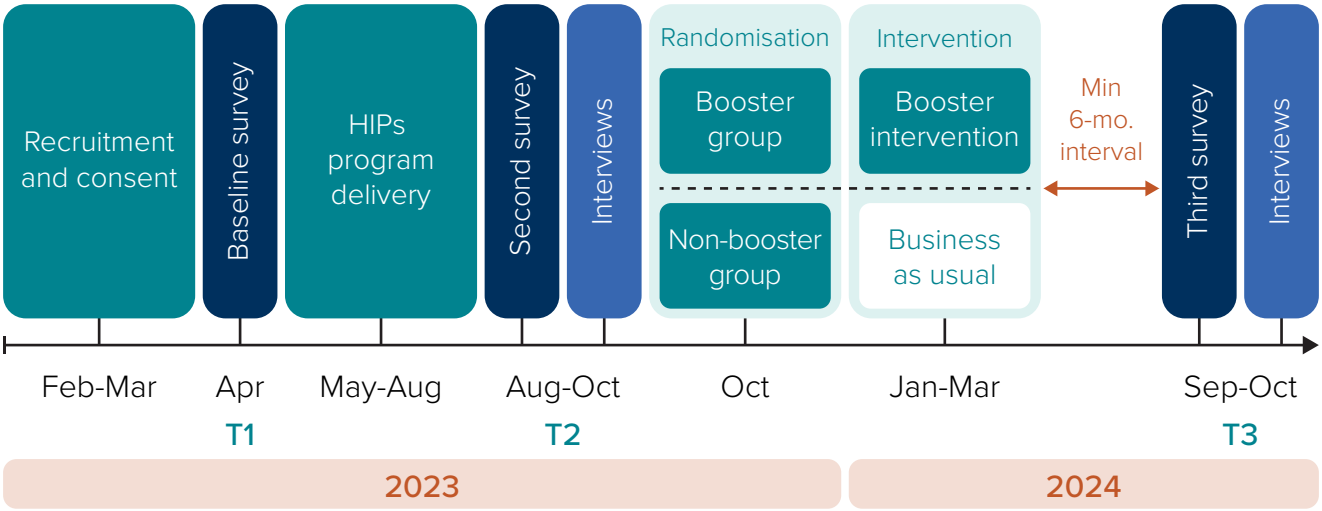
- consolidate and revisit prior learning
- reflect on progress and goals
- collaborate, discuss and problem-solve challenges implementing evidence-based practices with peers
- celebrate successes
- plan, set new goals and discuss next steps.

There were 3 primary data collection points throughout the lifetime of the project:

- a baseline measurement prior to participation in the HIPs program
- a measurement after completing HIPs and before randomisation and booster participation
- a final endline measurement 6 months after the booster program (one year after the conclusion of the HIPs program).

At each of these timepoints, data was collected from teachers and leaders through surveys and interviews. HIPs facilitators were also interviewed (Figure 1).

Figure 1: Study design and timeline



Note: All participating teachers and leaders were invited to fill out the 3 surveys; HIPs facilitators, school leaders and some teachers were invited to participate in interviews.

2.1.3. Sample and data sources

Schools: 17 schools in Western Australia participated in the research project, with 5 schools in the final Booster group allocation and 12 in the Non-Booster group (see Appendix A.2.1 for details on the randomisation and allocation process). All 17 schools were AISWA member schools. Two were primary schools, one was a secondary school and the remaining 14 were combined primary and secondary schools. One school was located in an outer regional setting, with the remaining 16 in major cities. Participating schools had a mean enrolment of 898 students (SD = 470, min = 187, max = 1670), and a mean Index of Community Socio-educational Advantage (ICSEA) of 1079.53 (SD = 46, min = 1003, max = 1175).

Teachers: From each school, 5 to 10 teachers were invited to participate in the research. To be eligible, teachers needed to be primary or secondary classroom teachers, be able to complete the HIPs program and all data collection surveys[2] and not have previously completed the HIPs program. An initial list of eligible candidates was provided by each leader, but individual consent was obtained from each teacher, and it was emphasised that participation was voluntary and they could withdraw at any time. A total of 107 teachers consented to participate in the research project, with baseline data obtained for 33 teachers in the Booster group and 69 in the Non-Booster group, and endline data obtained for 23 in the Booster group and 47 in the Non-Booster group. Teachers completed surveys that collected information on demographics and experience, school culture, confidence and use of evidence-based practices, perceived student outcomes, and perceptions of the HIPs program and booster intervention (if applicable). A subset of teachers additionally participated in semi-structured interviews to provide qualitative data on their experience of the HIPs program and booster, as well as perceived barriers and enablers to changing practice (13 from the Booster group, 7 from the Non-Booster group). In addition to teaching responsibilities, 29 teachers had leadership roles, with 31 teaching exclusively primary school classes, 63 teaching secondary and 2 teaching classes across both levels. On average, teachers in the sample had taught for 12.5 years and had been at their current school for roughly half of that time (see Appendix A.5.1 and Table 7 for details).

2 Teachers that completed the HIPs program but did not complete all data collection surveys could still participate in the booster intervention.

Leaders: One leader from each of the 17 participating schools participated in the research. These were predominantly deputy and assistant principals. Leaders did not directly participate in the professional learning with their staff, but completed surveys collecting information on their perceptions of the HIPs program, booster (if applicable) and the barriers and enablers to changing practice based on their experiences as a leader during the project. All participating leaders were invited to participate in interviews after their staff completed the HIPs program and at the conclusion of the study, with 11 completing at least one interview. On average, leaders in the sample had worked in education for 25.5 years and had been at their current school for 8.7 years.

AISWA HIPs facilitators: The HIPs program was delivered to participating schools by 8 AISWA facilitators, all of whom participated in interviews and collected fidelity and attendance data throughout the project to ensure that the professional learning was delivered and attended as intended. Each school had a single AISWA facilitator responsible for conducting the classroom observations with coaching and feedback. One of the 8 facilitators delivered the booster intervention across all of the 5 Booster group schools.[3]

2.2. Outcomes and reporting measures

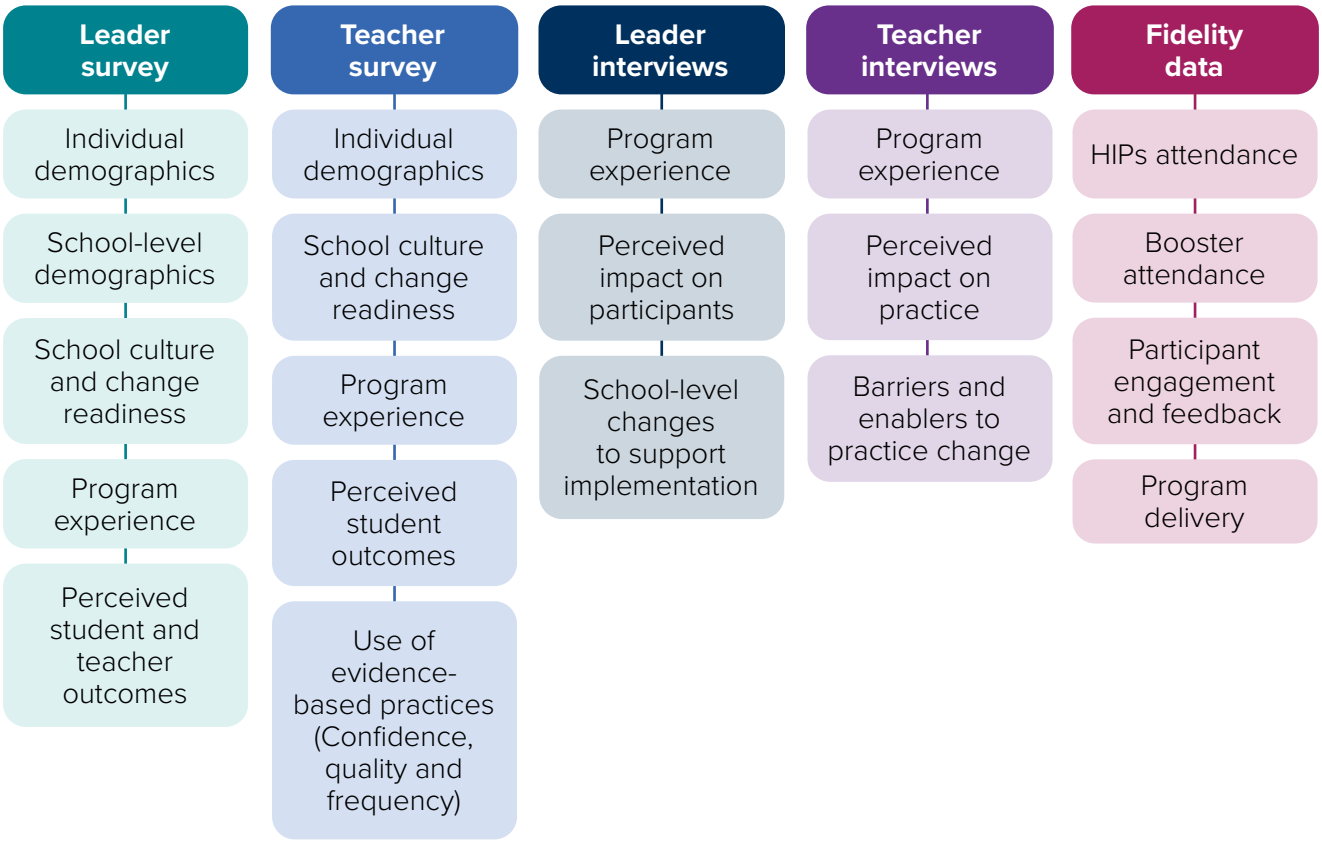
Data for quantitative analyses was collected through self-report measures within the surveys, and data for qualitative analyses were collected through semi-structured interviews and open-ended survey questions (Figure 2). The primary outcome measures used in the study were teacher-level outcomes focused on quality of use of evidence-based practices and confidence in using evidence-based practices, as well as perceived student outcomes (see Appendix A.4.1 for details on primary outcome measures).

Teachers' quality of use of evidence-based practices was captured using a 21-item rubric developed by AERO to assess the quality of explicit instruction and formative assessment practices. Within the surveys, the rubric is presented as a series of multiple-choice questions describing different approaches for a specific teaching technique or strategy. For each question, teachers select the description that most closely aligns with their practice. This measure has been validated and found to be sufficiently accurate to assess aggregated changes in teaching practice as a dependent variable (Robertson et al., 2024; see Appendix C for details).[4] Confidence was measured using 6 Likert scale questions derived from the same rubric, with a 1 to 5 response for each item (Not at all confident – Very confident). Perceived student outcomes were assessed using 3 Likert scale questions, with a 1 to 5 response assessing the proportion of students that 1) show active participation in their learning, 2) show confidence in their learning, and 3) are performing at the expected standard of achievement for their year level.

3 This booster facilitator was the same one used for the initial HIPs program for 3 of the 5 schools in the Booster group. This 'facilitator consistency' was used as a covariate in booster-only analyses but was not found to be significant for any dependent variable (see Appendix B.1).

4 This measure is used to assess whether a practice change has occurred, not to quantify the change.

Figure 2: Data collection overview



2.3. Program fidelity

The HIPs program was delivered across the 17 participating schools over a 3-month period by 8 experienced AISWA facilitators. These facilitators collected fidelity data on participating teachers' workshop session attendance and classroom observation completion. Of the 107 teachers who consented to participate, 82 to 93% attended the 3 workshops and 88% completed one or more classroom observation and feedback sessions. Overall, teachers and leaders reported very positively on their experience of the HIPs program (averaging 4.62/5 across a 5-item measure; see Appendix A.5.1).

The booster program was delivered separately to each of the 5 Booster-group schools by a single experienced AISWA facilitator and comprised 2 webinars approximately 2 weeks apart. In line with prior research (see the Elements of successful boosters), parts of the booster were designed to be responsive to participating teachers, particularly those relating to sharing experiences in making changes and implementing evidence-based practices.[5] Because of this responsiveness, the experience of the booster may have differed slightly between teachers. For example, different amounts of time could have been spent discussing goal setting and action planning across schools, and individual teachers may have been more likely to share their experiences or seek feedback. This could have affected the extent to which the initially planned 'balanced design' of the booster was achieved for each teacher, especially given the time constraints of the booster intervention (2x 90-minute sessions).

5 Research shows that individualisation is an important component of successful boosters. See Elements of successful boosters.

Facilitators reported some slight challenges in program delivery, including technology issues, content overload, and coordination issues at certain schools, but indicated that overall delivery was highly successful and very close to delivery as planned.

Of the 33 teachers eligible to participate in the booster, 77% attended the first booster webinar and 70% attended the second. As with the HIPs program, the booster intervention was well received by teachers (averaging 4.38/5 on the same 5-item measure; see Appendix A.5.1). Teachers reported that there had been ample opportunity to engage in most of the intended components of the booster, although there was some appetite for a greater focus on reflection and planning around the use of evidence-based practices (see Appendix A.4.4). Of the teachers that attended the booster, 82% engaged in at least one unfacilitated peer support activity outside of the webinars (e.g., lesson planning, mentoring or classroom observations), with 60% engaging in 3 or more of these activities.

3. Key findings

This section provides an overview of the study's main findings, with results from statistical tests interpreted alongside participants' responses from interviews and open-ended questions. Findings presented in this section are those that relate most directly to the research aims. See Appendix B for further details on conducted analyses.

3.1. Teacher outcomes improved over the course of the study, but there was no additional benefit from the booster

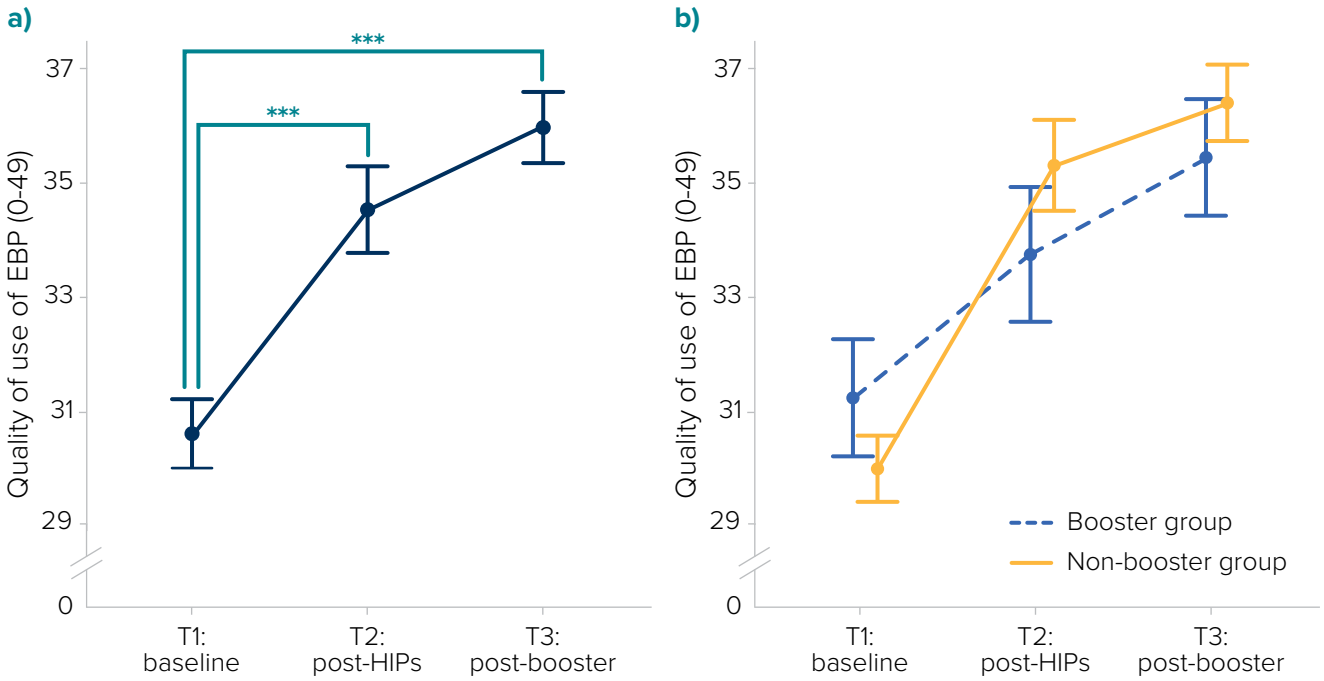
This study examined whether scalable booster interventions could act as an effective follow-up to a professional learning program. We saw improved teacher outcomes following participation in the HIPs program, indicating that we were set up to effectively test the contribution of the booster, but there was no evidence of any benefit for teachers who received the booster intervention compared to those who did not. Teachers in the Booster group consistently reported that the booster sessions had been helpful, but the 2 groups remained comparable on each measure throughout the study. In both groups, teachers' reported confidence to use evidence-based practices and quality of use of evidence-based practices significantly increased from baseline scores, alongside a positive, although non-significant trend for perceived student outcomes. Given the significant change between baseline and the data collection timepoint immediately after the initial professional learning delivery, it is likely that the HIPs program drove the improvement, although without a control group that did not participate in HIPs, we cannot attribute changes directly to the HIPs program. Overall, teacher outcomes improved over the course of the study, but the booster did not meaningfully contribute to these outcomes.

3.1.1. Quality of use of evidence-based practices

Compared to baseline, teachers participating in the study reported overall improved quality of use of evidence-based practices at both T2 and T3 (after the HIPs program and at the end of the study, respectively; see Figure 3a).[6] These improvements were highly significant, and although there was an additional trend towards T3 scores being higher than T2 (suggesting that practice quality may have continued to improve, albeit at a slower rate after the professional learning intervention), this difference was not significant for any group.[7]

Teachers who received the professional learning booster did not differ in their quality of use of evidence-based practice from teachers who received the initial professional learning only.[8] To isolate the period in which we would expect to see the impact of the booster, we compared the change in the 2 groups from T2 to T3 (after completing the HIPs program and 6 months after the booster intervention had been delivered). If the booster had been beneficial, we would expect to see a significant difference from this comparison, indicating that there had been an additional gain or less decay in the Booster group relative to the Non-Booster group. This was not the case. There was no difference in how the groups' reported quality of evidence-based practice use changed between T2 and T3.[9] The 2 groups remained very comparable throughout the study, showing a similar improvement after participating in the HIPs program and over the following year until the end of the study (see Figure 3b).

Figure 3: Quality of use of evidence-based practices over time

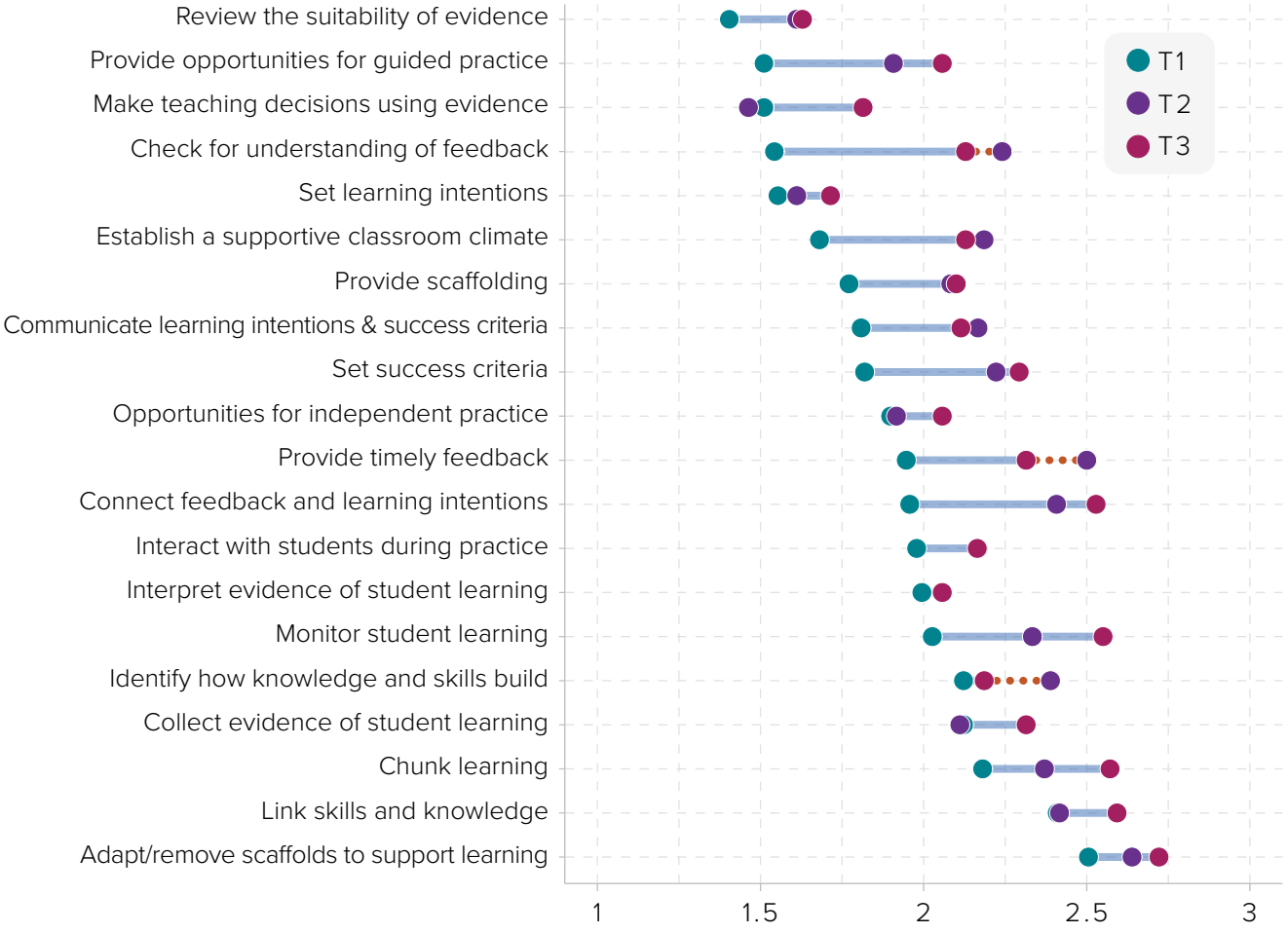


Note: Panel (a) shows quality of use of evidence-based practice across both groups combined, with panel (b) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line. Error bars represent ± 1 standard error.

6 **T1 to T2:** $t = 4.41$, Holm-Bonferroni corrected p-value < 0.001; **T1 to T3:** $t = 6.50$, corrected p-value < 0.001.
7 **T2 to T3:** Overall, Booster, Non-Booster $t = 1.55, 1.12, 1.11$, respectively; all corrected p-values > 0.05.
8 **Effect of group allocation:** $t = 0.58$, $p > 0.05$.
9 **T2 to T3 group interaction:** $t = 0.31$, $p > 0.05$.

We also examined use of individual evidence-based strategies that make up the evidence-based practice at the 3 timepoints. There was a general trend for improvement over time, although there was substantial variation between strategies. The most consistent improvements were shown for use of guided practice and feedback (see Figure 4 and Appendix B.1). These changes were fairly consistent across the whole sample, with no apparent differences between groups.

Figure 4: Quality of use of evidence-based practice strategies



Note: Raw scores for individual rubric items vary from 0 to 2 or 0 to 3, but in this figure, all scores have been normed to a 0 to 3 range for comparison. Higher scores represent better quality. Raw scores were used for all analyses and other figures. Solid bars in blue represent an improvement compared to the previous timepoint, dotted lines in orange represent a decrease.

Teachers and school leaders who participated in interviews were generally very positive about the impact of the HIPs program on teaching practice. Teachers spoke most about the impact at T2, after completing the initial HIPs program, frequently referring to particular strategies such as daily review, engagement norms, questioning and feedback when reflecting on the impact of the program on their practice. For example:

> I think before the HIPs program, I was rushing through the think-pair-share process. Now my students really understand that there are three steps, which has really benefited them and also my teaching practice.
>
> — Teacher, T2

> Having that input and being reminded of the theory or the evidence behind why feedback's so powerful has had that influence of, 'I need to step up here, this is an area where I'm almost losing traction.'
>
> — Teacher, T2

This emphasis was also echoed by school leaders and facilitators at T2 after the professional learning program:

> The main change that's coming through and [that] seems to be the thing that people have jumped on more is with how we do review in our classroom.
>
> — Teacher, T2

> I think where [teachers] have had an opportunity to focus on a particular area, I have seen improvements in practice in relation to that area. Probably the most common one is daily review, because that's a core part of the program in which the first observation is based. There's some improvements in practice that I've noticed from teachers' participation in the program.
>
> — Facilitator, T2

Although there was no detectible benefit from the booster intervention for quantitative outcomes, teachers generally felt positive about the impact that reviewing, clarifying and consolidating content during the booster had on their practice. Teachers commented that:

> It was very useful for consolidation because I had had a go, and was doing it for a while without any feedback, and then we came back during the booster, [it helped to] figure out where I am on the journey of applying it all. That was really helpful.
>
> — Teacher, T3

> I think the booster was quite well-timed. By the time the booster came around, most people had faded off [with using the practices], and it was a nice, timely reminder.
>
> — Teacher, T3

Teachers who participated in the booster also specifically highlighted the intervention's potential to drive practice change in having them re-engage with key content:

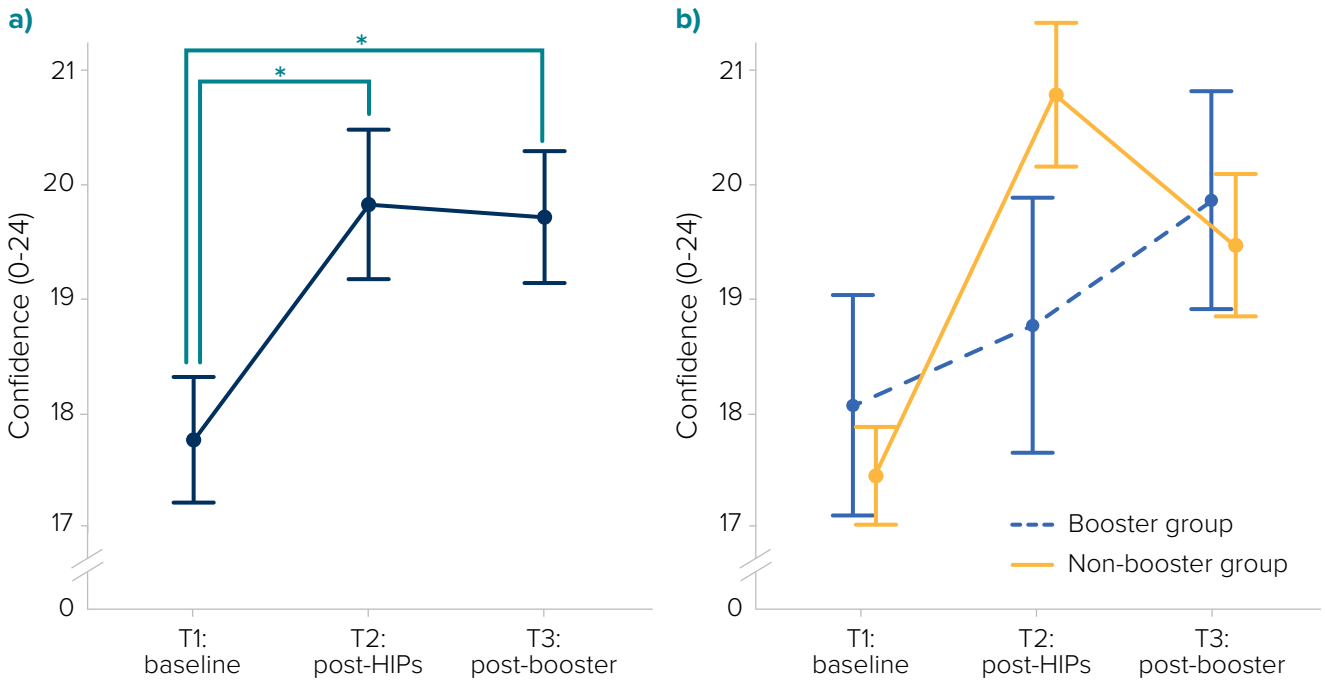
> I think the booster kept the conversation going. And it also keeps you accountable. I did say I was going to do daily reviews, so I'd better actually do that.
>
> — Teacher, T3

> I feel like the boosters provided that impetus or that drive to continue.
>
> — Teacher, T3

3.1.2. Confidence to use evidence-based practices

Compared to baseline, teachers participating in the study reported improved confidence in using evidence-based practices at T2 and T3 (after the HIPs program and at the end of the study, respectively; see Figure 5a),[10] but there was no difference between the Booster and Non-Booster groups.[11] There was also no significant difference between scores at T2 and T3 for any group,[12] suggesting that there was no meaningful improvement following the booster and no meaningful decay over the year following the HIPs program for teachers in the Non-Booster group. Although there was a trend towards a Booster group advantage from T2 to T3 (see Figure 5b), it was not statistically significant.[13] Additionally, this trend may have been driven more by the higher scores in the Non-Booster group at T2 than by any meaningful benefit from booster participation, as scores at the beginning and end of the study were very similar. At the level of the individual items used to assess confidence, each component showed higher mean ratings at T2 and T3 compared to baseline, although there were no consistent differences between scores at these 2 points (see Appendix B.1.2).

Figure 5: Confidence to use evidence-based practices over time



Note: Panel (a) shows confidence to use evidence-based practices across both groups combined, with panel b) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line in yellow. Error bars represent ± 1 standard error.

10 **T1 to T2:** t = 2.36, Holm-Bonferroni corrected p-value = 0.040; **T1 to T3:** t = 2.44, corrected p-value = 0.033.
11 **Effect of group allocation:** t = 0.47, $p > 0.05$.
12 **T2 to T3:** Overall, Booster, Non-Booster t = 0.13, 0.75, -1.40, respectively; all corrected p-values > 0.05.
13 **T2 to T3 group interaction:** t = 1.39, $p > 0.05$.

In interviews, teachers indicated that participation in the HIPs program had increased their confidence around teaching practice, aligning with the quantitative survey measures:

> We had some really good conversations that came out of the course that have definitely strengthened my confidence and understanding around pedagogy.
>
> — Teacher, T2

This sentiment was particularly evident at T2, after the HIPs program, and was also echoed by leaders at participating schools, supporting the notion that program participation improved teacher confidence:

> It's just given them that little turbo charge where they've actually understood the theory a little bit more, they've had some feedback on how they're using it, and they've massively grown in confidence by doing that.
>
> — Leader, T2

One leader also explained the additional effect the booster sessions had on their staff:

> There are some small changes I've seen. I would say for one of the participants, it was more of a confidence in the approach they took in the classroom.
>
> — Leader, T3

3.1.3. Perceived student outcomes

Teachers were asked about their perceptions regarding 3 student-level outcomes:

- student engagement
- student confidence
- whether students were performing at the expected standard of achievement for their year level.

Over the course of the study, there was a consistent trend towards improved perceived student outcomes (as reported by teachers; see Figure 6a), but the improvement was not statistically significant at any timepoint.[14] Nevertheless, the trend was echoed at the level of the individual items, with each component showing large increases relative to baseline (see Appendix B.1.3). Given the overall trend, and the fact that perceived student outcomes were scored using a scale with only 3 items and limited response options, it is plausible that we lacked the sensitivity to detect a real improvement in perceived student outcomes, although any such benefit would likely be minor.

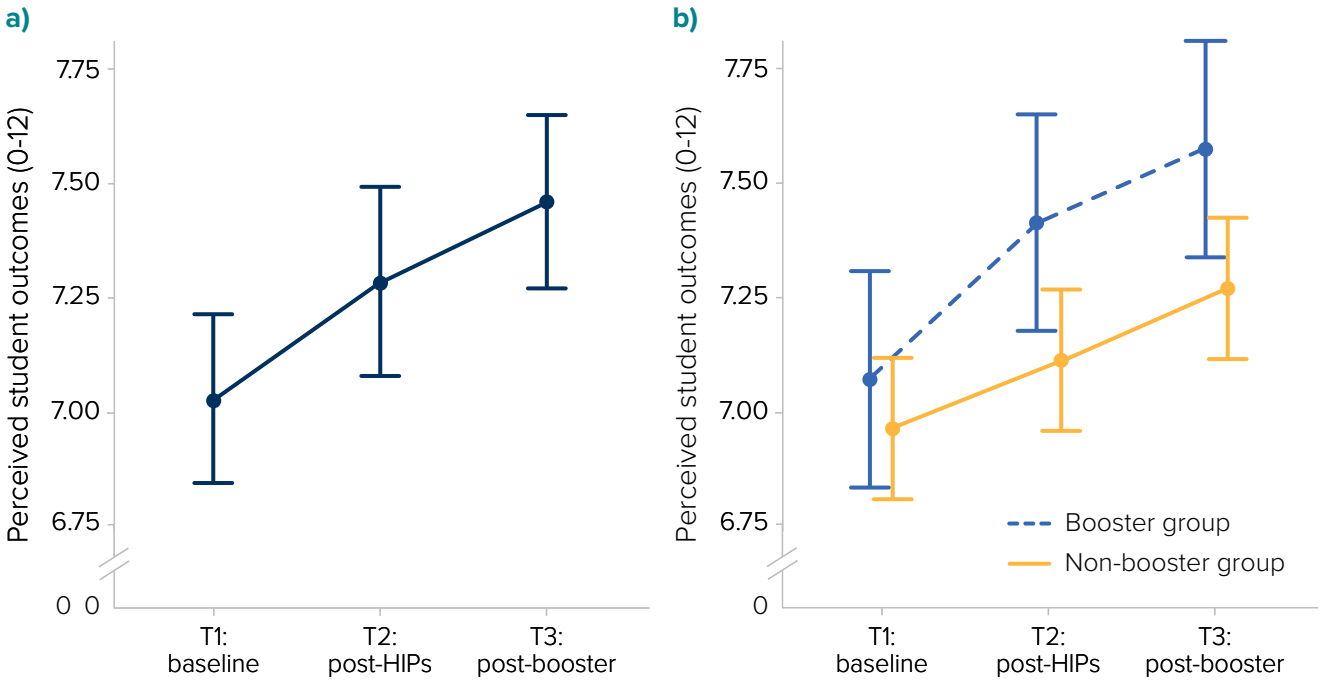
14 **T1 to T2:** $t = 1.03$; **T1 to T3:** $t = 1.85$; **T2 to T3:** $t = 0.71$, all corrected p-values > 0.05.

One leader, however, acknowledged the direct relationship between changes in teaching practice and the positive results teachers saw at the end of the year:

> When I sat down with one of the staff members and reflected on their end-of-year results, we noticed his results were really positive. And I said to him, 'What's the tip?' And he said, 'Oh, HIPs, absolutely', which was really interesting. He said he'd changed his teaching. He changed his teaching style, and he believes that that contributed to the improvement in results at the end of the year.
>
> — Leader, T3

There were no meaningful differences in perceived student outcomes between the 2 groups. They remained similar across the course of the study, and although there was a trend towards a slight advantage in the Booster group, this began before participating in the booster sessions and was not significant at any timepoint (see Figure 6b).[15] Importantly, there were no differences in how outcomes changed following the booster intervention. Teachers reported the same trend towards improvement in perceived student outcomes from T2 to T3, whether or not they received the professional learning booster.[16]

Figure 6: Perceived student outcomes over time



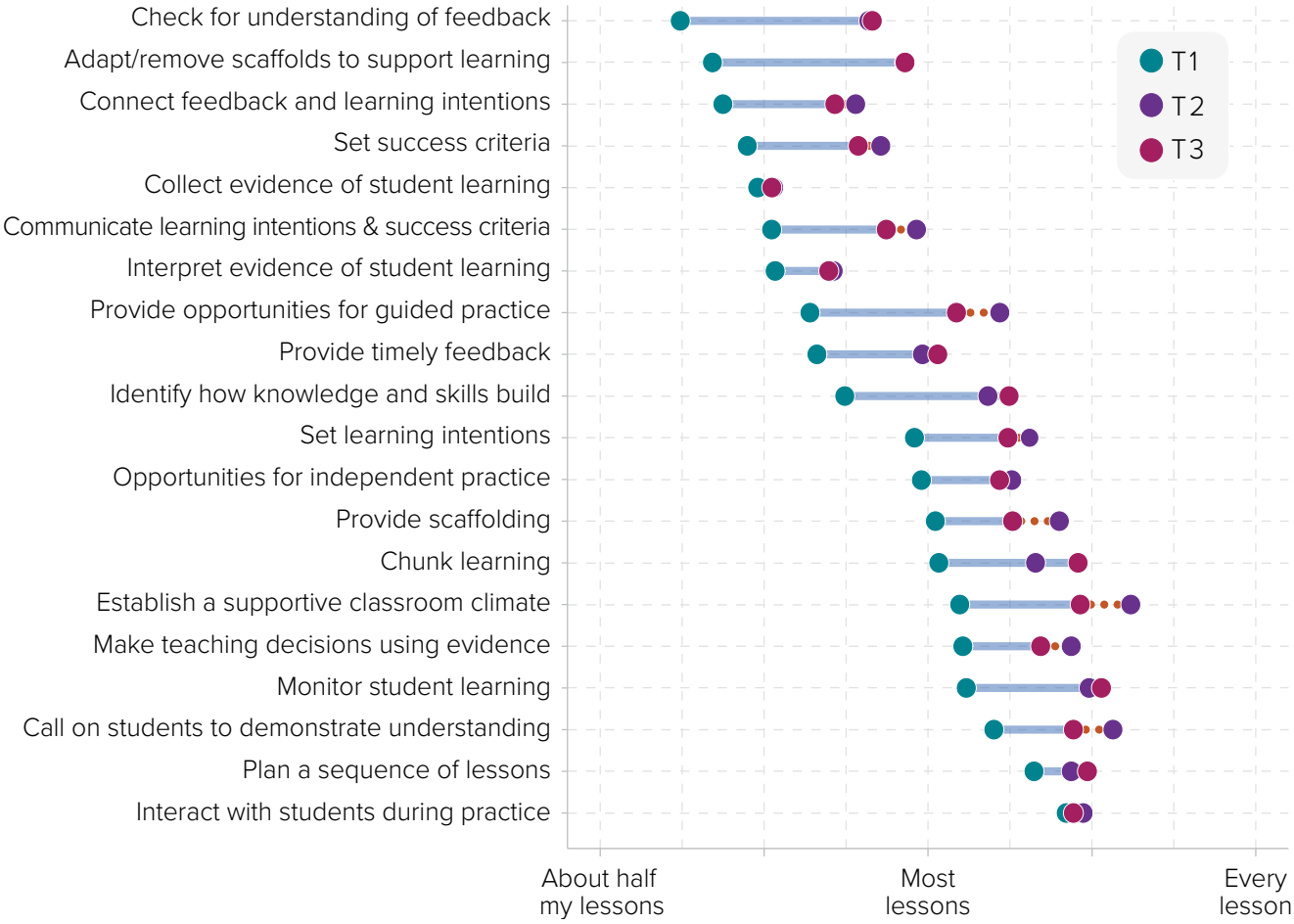
Note: Panel (a) shows perceived student outcomes across both groups combined, with panel (b) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line. Error bars represent ± 1 standard error.

15 **Effect of group allocation:** $t = 1.09$, $p > 0.05$.
16 **T2 to T3 group interaction:** $t = 0.03$, $p > 0.05$.

3.1.4. Frequency of use of evidence-based strategies

We collected data on teachers' frequency of use for a range of evidence-based strategies to describe what teachers were already doing and explore changes in use of the strategies (see Figure 7). This frequency data aligned with the data on quality of practice use, with the greatest increases seen around guided practice and feedback, and no apparent differences between groups.[17]

Figure 7: Frequency of use of evidence-based strategies over time



Note: Teachers rated the frequency of their use of each strategy, selecting from 'Never', 'Some lessons', 'About half my lessons', 'Most lessons' and 'Every lesson'. Solid bars in blue represent an increase compared to the previous timepoint, dotted lines in orange represent a decrease.

17 Due to the non-linear response scale used for frequency, and variability in what could be considered desirable for each strategy, this data was treated as purely descriptive, and was not aggregated for use in any quantitative analyses.

3.2. Components of the booster were highly valued, but time and flexibility were barriers to success

3.2.1. Collaboration

Over the course of the study, the themes of collegial collaboration, connection and sharing became evident from both teachers and school leaders when reflecting on the impact of the professional learning booster, all closely aligned to the effective mechanism of practical social support. These were seen as additional outcomes of participation, and for the majority of participants, they were also seen as enablers to practice change.

At the end of the study, one teacher commented:

> The fact that we're there to support and help each other has fostered that ability to work with these strategies and to implement them in a way that's positive.
>
> — Teacher, T3

School leaders were also focused on the broader benefits of bringing together staff teams as part of the professional learning. These benefits included creating a shared language across their school and facilitating conversations about improvement:

> There was good, robust discussion that happened … what we're trying to achieve in a group like this is that staff are regularly looking to seek ways to improve and have conversations with their peers.
>
> — Leader, T2

Face-to-face sessions were seen as particularly useful for allowing collaboration with peers and contextualisation of content. Participants emphasised the value in being able to discuss the content with their peers during the booster sessions, once they had had an opportunity to attempt implementation themselves. For example:

> Completing the course with colleagues from my school worked well. We all had the same context and this made the discussions that we had more valuable. It was very worthwhile hearing the perspectives of my colleagues.
>
> — Teacher, T2

> I think just being able to share what you were doing in your classroom. Because everyone did it differently. But also, people to share where they struggled with it. So, not only were you able to pick out the victories, but you were also able to discuss those ways that you could improve it further.
>
> — Teacher, T3

> I do know that the booster sessions at the beginning of the year were a wonderful chance to reconnect. My understanding is that several of those staff reconnected on a more informal level, maybe one on one with certain people to share some practices and have some conversations about, 'Hey, what do you do in this situation?' Or 'Can I come in and see the beginning of one of your lessons?' So that collaborative piece has been something that I know has happened.
>
> — Leader, T3

> The [booster] to me was better than all the other stuff that we did but part of that was we'd actually had time to try [the practices]. We actually got more time to discuss the things that we'd tried and what we were doing.
>
> — Teacher, T3

> I did get the sense that participants valued hearing one another's perspectives, and how they implemented practices within their area of specialty. I think that sharing of practices was really valuable and I received positive feedback from teachers.
>
> — Facilitator, T3

3.2.2. Content review and momentum

In addition to the emphasis on the value of the boosters in allowing teachers to reconnect after trying to implement practices themselves, teachers appreciated the opportunity to review the content and maintain momentum, in line with the core booster mechanism of revisiting prior content. Several teachers noted that the timing was particularly helpful, with the booster intervention arriving at a point where momentum might otherwise be waning:

> I think it helped to remind and get re-inspired to keep going.
>
> — Teacher, T3

> I think the booster was really good for trying to get out of that rut of, 'Oh, I'll just go back to what I've always been doing', and to actually make some changes. It was super helpful.
>
> — Teacher, T3

> I think having an opportunity for teachers to revisit prior learning is wonderful. I feel like it could be enhanced by having the teachers co-construct what that might look like, so it can reflect their needs at that particular time in that particular context, and also based on what's possible with school leadership.
>
> — Facilitator, T3

3.2.3. Staffing and time

Although participants were very positive about the booster intervention in interviews, they did highlight some key challenges that could have contributed to the lack of meaningful improvements. Staff consistency and time were both cited as challenges for enacting change following the booster. One teacher commented that:

> We've had a lot of staff change since we did the program, even since the boosters. We've got some new staff who are really motivated and inspiring, which is great, and I really wish they'd been here to do that as well. And some of the staff who were doing it have either left us, or completely disengaged with teaching and have no intention of implementing anything.
>
> — Teacher, T3

However, the most prevalent issue was time:

> The day-to-day busyness of teachers means that once it's taught, it gets put to the side and they move to the next one. And I think that's a bit of a gap that we need to work on.
>
> — Leader, T3

> Other than time together and to implement it, I don't actually feel like there are any barriers, really.
>
> — Leader, T3

> I'm not finding the time to implement something or try something different or even prepare something new. I just have to run with the tried and tested.
>
> — Teacher, T3

> I think the business of schooling just took over.
>
> — Leader, T3

Despite the brief nature of the booster, time remained a consistent barrier to changing practice because the time for implementation and experimentation outside the booster sessions remained as difficult to find as it would be alongside any professional learning intervention. This indicates that while it may be relatively easy to manage time and staffing *for participation* in a short booster intervention, it is still challenging to provide sufficient opportunity for teachers to work on changing their practice outside of those sessions. Schools may be better off focusing their time and resources on the successful implementation of professional learning programs, rather than engaging in light-touch follow-up interventions.

3.2.4. Flexibility

While allowing time for tailored discussions was an important aspect of the booster intervention, including it meant that the delivery of effective mechanisms of professional learning and the balanced, scalable design of the booster may have differed from school to school.

> There was enough flexibility within the structure of the booster to be really responsive, and allow the conversation to expand a little bit here and there in whatever school and context that I was working with. I think that that was a real strength of that booster.
>
> — Facilitator, T3

Most teachers who participated in boosters indicated that they had ample opportunity to revisit content and reflect on their practice, but there was some variability. About two-thirds of teachers indicated they had enough time to revisit goals and share with their colleagues, but only half of the sample felt there was as much time as they would have liked for action planning, and some teachers also reported not having enough time to troubleshoot problems and share learnings with their colleagues (see Appendix A.4.4). These reports varied between teachers from the same school (and, therefore, the same session), suggesting that there may not have been enough opportunity to account for individual needs and preferences within the constraints of the light-touch booster.

3.3. Limitations

3.3.1. Booster sample

The Booster group comprised a smaller proportion of the total sample than originally intended due to 3 schools declining to participate in the booster sessions (our final sample contained 12 Non-Booster schools and only 5 Booster schools). This reduced the effective sample size of the intervention group and may have also affected the representativeness of the group, as schools that differed in systematic ways may have been more likely to refuse the booster. For instance, facing staffing or resource issues during project delivery might have made it more difficult to implement practice changes following the initial program, and also made schools more likely to refuse the booster. This could, in turn, lead to an overestimation of the benefits of the booster, as schools that did not face any of these challenges would be overrepresented in the Booster group. To evaluate this possibility, supplementary analyses were run to compare the schools that refused the booster to the rest of the sample, and to test whether the overall pattern of results differed with these schools included in the Booster group as originally allocated (see Appendix A.5 and Appendix B.3). These analyses indicated that the booster-refused schools were comparable to the rest of the sample, both before and after the intervention, and that the same overall pattern of results was obtained from analyses using their original allocation to the Booster group. Further, the most likely issue arising from this limitation would have been an apparent group difference leading to an overestimation of the benefits of the booster, but we did not find differences between study groups at any point. Overall, booster-refusal is unlikely to have substantially influenced the pattern of results, although it would have reduced statistical power of the analyses (see Section 3.3.4), and may indicate that even light-touch booster interventions would be difficult for schools to integrate in addition to sustained professional learning programs.

3.3.2. Unaccounted-for concurrent professional learning

It is possible that group allocation in this project may have influenced how likely participating schools were to engage in concurrent professional learning, creating an imbalance that could have masked benefits of the booster intervention. Schools in the Non-Booster group may have been more likely to participate in other professional learning during the study period, whereas Booster group schools were already committed to more professional learning as part of the project and may have had less capacity or felt less need to engage in more. Any additional professional learning undertaken by the Non-Booster schools during the project could have improved their outcomes to a similar degree to the booster intervention, meaning that a benefit of the booster would not be detectable as a difference between groups.

School leaders in both groups were asked whether their staff had engaged in other professional learning during the study, with all respondents indicating that their staff had participated in other professional learning during the research project. While this suggests that concurrent professional learning was common across both groups, responses did not clarify whether the staff involved were those that were directly participating in this research or other teachers at their school, and responses were not obtained from all participating schools, meaning we cannot completely rule out the possibility that teachers in the Non-Booster group participated in concurrent professional learning. The intensity of the concurrent professional learning was also not recorded, which could further contribute to changes in those schools, although given the time constraints reported by participants, it is unlikely that these would be the sorts of sustained interventions that are likely to be effective.

3.3.3. Measurement sensitivity

It is possible that the measures used in this study were not sensitive enough to detect subtle improvements in teaching practice that may have resulted from the booster. While Likert-type measures such as those used to assess confidence in this study are commonplace in research, the rubrics used to assess teaching practice are novel and were developed specifically for studies like this one. However, all these measures were able to detect overall improvements across both groups across the duration of the project, and external validation work (see Appendix C) has provided evidence for the utility of AERO's formative assessment and explicit instruction rubrics, indicating that they are fit for use assessing aggregated teacher-level changes. This suggests that the instruments should be able to capture improvements, and the lack of non-significant trends toward booster benefits is better explained by a lack of meaningful intervention effects than a lack of measurement sensitivity.

3.3.4. Statistical power and sample size

The final sample of schools in the study was quite small due to the constraints on recruitment, limiting our power to detect effects. This was then compounded by the smaller proportion of schools ultimately receiving the booster intervention (see Section 3.3.1 and Appendix A.2.1). Despite being close to our initial recruitment goal based on power calculations (see Appendix 2.1.2), our Booster group ended up substantially smaller than anticipated and we were not powered to detect small or moderate intervention effects. We paid particular attention to trends across groups in both the quantitative and qualitative data. Overall, it appears that the reduced statistical power is not likely to have masked a meaningful benefit of the booster, as there were no consistent indications of a Booster group advantage in any form of data, whereas group-independent effects over time were clearly demonstrated.

3.3.5. Generalisability

While this study found no evidence of effectiveness from this specific booster intervention, we cannot rule out the potential value of other booster designs. It is possible that alternative booster formats (varying structure, timing, content or delivery method) could result in meaningful improvements. However, the booster design that we tested in this study was based on best practices identified in existing research, given the constraint of maintaining the scalability of the intervention. It is plausible that a booster intervention that is more intensive, individualised or context-specific would be more effective, but this might come at the cost of scalability. Within a brief intervention, there would be little opportunity to incorporate more effective mechanisms of professional learning. As an additional consideration, this study was mostly conducted in fairly large, combined primary and secondary schools located in major cities, and the findings may not extend to distinct educational contexts.

4. Implications and next steps

4.1. Quality professional learning is required to improve evidence-based teaching practice

Following the HIPs program, teachers in this study showed an increase in their confidence to use evidence-based practices and reported quality of use of evidence-based practices. We found no further benefits for teachers who received the scalable booster intervention on top of the initial professional learning program. These results align with existing research on effective professional learning, indicating that for interventions to improve teaching practice, they need to be sustained and responsive, with sufficient opportunities to practice skills and embed them in context, in addition to incorporating effective mechanisms (Sims et al., 2021; Wei et al., 2010). An intervention such as the booster, that includes these mechanisms but is not sufficiently sustained and responsive, will not improve teacher outcomes.

The timing and improvements seen over the course of this study for both groups indicate that the changes detected were likely driven by the HIPs program, our initial professional learning intervention. Given that this study lacked a comparison group that did not participate in HIPs, we cannot draw causal conclusions about the efficacy of the program, however, the evidence is suggestive.

The HIPs program utilised many of the effective mechanisms of professional learning identified in the EEF's systematic review and meta-analysis (Sims et al., 2021), and would qualify as a balanced design (see Appendix A.3.1). Similarly, in accordance with Darling-Hammond et al.'s (2017) review of effective professional development programs, the HIPs program was content-focused, incorporating active learning and collaboration, included expert support and feedback and was of a sustained duration. According to both taxonomies, the HIPs program conforms to expectations about what effective interventions entail, and the improved teacher outcomes support this interpretation.

In contrast, the scalable booster intervention delivered in this study had no significant impact on teacher outcomes despite being based on research and including many of the effective mechanisms of professional learning. The booster intervention was also intended as a balanced design, touching on the 4 requisite categories of insight (revisiting prior learning), goals (goal setting and reinforcement), techniques (practical social support and feedback) and embedding practice (action planning and self-monitoring). However, the short, light-touch nature of the booster may have meant that even though many of the effective mechanisms of professional learning were included in the design, they could not be sufficiently leveraged to improve teaching practice. Another possibility is that because the booster was designed to be responsive to teacher needs, it may not have included all of those mechanisms for all teachers. The booster was designed to be scalable and minimally burdensome, which limited its length, depth and comprehensiveness compared to the initial training, and these constraints may have undermined its utility.

4.2. Promising booster components could be implemented elsewhere

Although the booster intervention in this study did not improve teacher outcomes, participants consistently reported that it had been useful and highlighted specific components they found especially valuable. These components were aligned with the effective mechanisms of professional learning and research recommendations for follow-up interventions and sustaining improvements. Although they were not effective within a scalable booster, they could be leveraged in other contexts as a follow-up to professional learning without the same constraints.

4.2.1. Collaborative support

In line with prior research and the mechanism of practical social support, the booster intervention in this study was designed so that multiple teaching staff from each school could attend together (Crowther & Cannon, 2002; Rusby et al., 2013). Across interviews and open-ended survey questions, teachers and leaders consistently reported that the booster sessions provided valuable opportunities to work with their peers to clarify their understanding of content. The collaborative follow-up sessions were also reported to create accountability that helped to sustain momentum after the initial program had finished. These elements are aligned with existing evidence on the mechanisms of effective professional learning (Darling-Hammond et al., 2017; Sims et al., 2021) and effective follow-up interventions (Edwards Groves & Rönnerman, 2013; Elias, 2010; Prenger et al., 2017).

While collaborative support could be incorporated at the end of any professional learning program, it could also be encouraged directly by school leaders to maximise momentum after professional learning, perhaps within existing staff or team meetings. This would make use of the mechanisms of revisiting prior learning, context-specific repetition, feedback and practical social support over a longer period than a light-touch intervention could provide, and without needing to engage more externally facilitated professional learning.

4.2.2. Time to plan, troubleshoot and share learnings

In their feedback on the booster, some teachers reported not having enough time for action planning or to troubleshoot problems and share learnings with their colleagues in the booster sessions. Following professional learning, teachers may simply need more support than is feasible in a brief, scalable intervention to move beyond their initial gains or overcome specific barriers in applying new practices.

Time constraints for planning, troubleshooting and sharing are not unique to professional learning boosters and do not need to be delivered as part of an externally facilitated follow-up intervention. Schools that have the right conditions to participate in a booster intervention (e.g., a collaborative culture and supportive leadership) could embed these features within their own structures instead, potentially saving resources while increasing effectiveness. Redirecting the time and effort needed to deliver an externally facilitated booster into school-based collaborative structures may also address concerns raised by participants about the limited reach of follow-up sessions. Several leaders noted that it was difficult to translate the benefits of the professional learning to colleagues who had not directly participated. School-based follow-up approaches could support wider staff engagement, especially when the professional learning is aligned with broader school goals and delivered within existing structures or processes that reach more staff. This could extend the benefits of professional learning without the added cost or coordination required for additional program enrolments or externally facilitated follow-up interventions.

4.2.3. Addressing diverse teacher needs

Follow-up professional learning interventions are more likely to be successful when they are tailored to individual needs and include more contextually dependent and specific instruction (Anshu, 2019; Durlak & DuPre, 2008; Miller et al., 2014). While delivering the booster, it was challenging to provide this tailoring and deliver a light-touch, scalable intervention that used a balance of effective mechanisms of professional learning (Sims et al., 2021). Implementing a follow-up to professional learning through existing collaboration structures in schools (such as regular team meetings) could make it easier to tailor the content to suit individual teachers and their specific contexts. In an externally facilitated intervention, time needs to be spent establishing the needs of each participant and the specifics of their classroom context that might make generalised processes less effective. This is not always possible within brief, scalable interventions such as the professional learning booster in this study. Teachers and leaders within a school will already have knowledge of these specifics and could be better placed for appropriate action planning and troubleshooting with their colleagues. Although responsive delivery of these elements is difficult to achieve in a scalable booster intervention, following professional learning with structured collaboration within schools could help address individual needs and sustain practice improvements.

> What I think could be done would be to work with middle leaders and to let them deliver a booster within their school context. And allow them to co-construct and curate the content with a facilitator who has a set of prerequisites for what the booster needs to contain, but with enough flexibility for them to be able to ensure that it is contextually relevant.
>
> — Facilitator, T3

4.3. Current evidence does not support the efficacy of light-touch professional learning

Research into the efficacy of professional learning interventions most often demonstrates short-term outcomes, and the question of how to best sustain practice improvements through professional learning is a priority for future research, particularly at scale (Fullan, 2006; Wei et al., 2010; Zehetmeier & Krainer, 2011). Accordingly, we tested whether a scalable booster intervention could be used as a follow-up to effective professional learning programs to improve long-term outcomes, but found no evidence of efficacy despite incorporating many effective mechanisms of professional learning identified in recent research. Although these mechanisms are associated with improved outcomes from professional learning, they were not effective within the booster, suggesting that these kinds of interventions may not fail due to a lack of effective mechanisms, but because of the constraints in keeping them light-touch and scalable.

While the specific booster intervention we tested was not effective, it is still possible that other light-touch interventions could be more successful. We found challenges in delivering a brief, scalable booster tailored to teacher needs and incorporating a balanced set of effective mechanisms of professional learning. A greater focus on fewer mechanisms could potentially be more effective, especially those related to peer support and check-ins, as these were identified as useful booster components in our study. However, the meta-analysis run by Sims et al. (2021) emphasised that interventions that include more mechanisms and qualify as balanced designs tend to be the most effective, making a focus on fewer mechanisms less promising. More evidence is needed to explore what works best to address various teacher and school needs and circumstances.

Overall, current evidence from this and other studies (Darling-Hammond et al., 2017; Garet et al., 2001; Lord et al., 2019a, 2019b; Weiss & Pasley, 2006) shows that light-touch, scalable professional learning interventions do not meaningfully enhance teaching practice, either on their own, or after a period of professional learning. Enhancing teaching practice and sustaining change takes time and resources, and so far, there are no demonstrated shortcuts to achieving this. It is possible that future research could identify a more effective form of scalable booster, but unless this evidence emerges, schools and systems should avoid light-touch interventions and focus on leveraging the effective mechanisms of professional learning through quality interventions that are sustained in duration.

Appendix A: Methods

A.1. Ethics and ethics approvals

The research was approved and conducted in accordance with the requirements set out in the National Statement on Ethical Conduct in Human Research 2007 (updated July 2018). Throughout the course of the project, human research ethics review, monitoring and oversight were carried out by Macquarie University (HREC approval number 44581).

A.2. Sample selection

Schools located in Western Australia were eligible to participate in the study, provided they met the following criteria:

- They were an AISWA member school.
- High-impact teaching was identified as a school focus area.
- They were able to nominate 5 to 10 classroom teachers to participate, a team leader willing to act as a liaison with the AISWA facilitator and one school leader who had oversight of professional learning.

Seventeen schools and 107 teachers consented to participate in the research project through the following recruitment process:

1. AISWA issued expression of interest (EOI) recruitment material to member schools through email and social media.
2. After receiving an EOI from each school and confirming their eligibility, AERO emailed the principal with a link to an online form where they provided school-level consent and details for a representative responsible for continued communication with the research teams and the nomination of other staff for participation.
3. Representatives were then contacted via email to nominate and provide contact details of teachers and a school leader to participate in the research.
4. Nominated staff members were provided with participant information sheets and consent forms to indicate their consent to participate in the research. Consent forms emphasised that participation in the research was voluntary, that they could withdraw from the research at any time and that any queries and concerns about the research should be directed to the research team.

A.2.1. Randomisation

This study was a 2-arm randomised trial, with randomisation at the school level. The 17 participating schools were randomly assigned to one of 2 study groups after the delivery of the initial professional learning program, approximately 6 months ahead of booster delivery.

The initial random allocation for this study was produced using permutation randomisation without stratification, yielding a list of 8 schools for the Booster group and 9 schools for the Non-Booster group. All schools were notified of their allocation after the conclusion of the HIPs program. From this allocation, 2 schools ultimately needed to be moved from the Booster arm to the Non-Booster arm.

For one of these schools, post-HIPs data collection was late, and there would not have been a sufficient interval before the booster intervention to test its efficacy as a follow-up intervention. For the other school, due to high staff turnover, there would have been too few individual teachers participating in the study by the time of the booster for the intervention to support group collaboration. These schools did not receive any aspect of the booster intervention, and as randomisation only occurred after the conclusion of the initial professional learning, their experience in the study up to this point was identical to the rest of the schools that were allocated to the Non-Booster group. To maintain an even allocation across groups, replacements for these 2 schools were randomly selected from the Non-Booster group. However, by this point, all schools had been notified of their initial allocation, and both selected replacement schools indicated that they would not be able to participate in the booster due to conflicting obligations and time constraints during the selected delivery period. Rather than inviting additional schools over increasingly late allocation rounds, and exacerbating the risks to group representativeness (see Section 3.3.1), we elected to continue with the uneven allocation and conduct a series of additional checks to ensure that schools that could not participate in the booster were comparable to the rest of the sample on observable characteristics and that the refusals had not biased our results (see Appendix A.5.2 and Appendix B.3).

Shortly before the delivery of the booster, one last school indicated that they would not be able to participate in the intervention due to time and staff constraints. Given the proximity to the intervention, we again decided not to attempt to reallocate a school from the Non-Booster group, and to conduct additional checks comparing this late booster refusal school to the rest of the sample (see Appendix A.5.2). As with the other schools that could not participate in the booster, this school had not received any aspect of the booster intervention that differentiated their experience from the rest of the Non-Booster group, but they had been aware of their allocation to the Booster group for longer. This raised a concern that the school would not be representative of a business as usual comparison group over this period. However, group balance checks suggested that this school was comparable to the rest of the sample on observable characteristics (see Appendix A.5.2), and all primary outcome analyses were repeated as intention-to-treat (ITT) analyses using the original group allocations (regardless of booster participation) with the overall pattern of results unchanged (see Appendix B.3). All primary outcomes reported in the main body of the report include data from all schools with group allocation representing participation in the booster intervention. This final allocation consisted of 5 Booster schools and 12 Non-Booster schools.

A.2.2. Power calculations

Prior to recruitment, power calculations were performed to estimate the number of schools required to detect an effect between the 2 intervention groups, using the quality of evidence-use rubric measure (i.e., between-group outcomes; the ability to detect the difference between receiving a booster intervention or not, as a follow-up to professional learning received by all teachers).

The tested model was a clustered hierarchical model, with teachers nested within schools. We calculated the required sample to detect an effect at the teacher level. Calculations were based on an 80% chance of successfully detecting a non-directional effect at an alpha-level of 0.05 ($\alpha = 0.05$, $\beta = 0.8$), with an 80% teacher retention rate and a conservative 0.65 test-retest correlation factored in.[18]

18 Based on independent evaluation (see Appendix C), the AERO rubric measure was calculated to have a WLE reliability of 0.85 overall, but test-retest reliability was not estimated. For longitudinal analysis, test-retest is the critical determinant, and we elected to use WLE only as a rough upper bound to inform the selection of the conservative 0.65 placeholder.

Based on a review of the professional learning literature, the average effect size for teacher-level instruction outcomes was estimated at d = 0.49-0.66,[19] alongside an Intraclass Correlation Coefficient of 0.07 (Dong et al., 2016; Gesel et al., 2021; Hattie, 2023; Kraft, 2020; Morina et al., 2025). As the booster intervention was a follow-up to another professional learning intervention, we expected a relatively weak effect and used the lower bound of this effect size range (0.49). These calculations yielded an estimated minimum sample of 20 schools, assuming an average of 6 teachers per school (5 being the minimum, 10 being the maximum). Thus, the present study aimed to achieve a sample of 10 schools per study group, with 5 to 10 participating teachers per school.

Power calculations were repeated at the end of the recruitment process, based on the size of the obtained sample. We calculated power based on a total of 107 teachers clustered within 17 schools (continuing to factor in a further 20% attrition on the obtained sample). These calculations indicated that the minimum detectible effect size (MDES) for the obtained sample was 0.52, suggesting we were likely slightly underpowered to detect an effect from the booster. Following the declining of the booster by some schools (see Appendix A.2.1), we conducted the power calculations again, based on the uneven proportion of teachers remaining in each group. These calculations yielded an MDES of 0.55 with booster-refused schools folded into the Non-Booster group, and an MDES of 0.58 with the booster-refused schools removed from the sample. Although these estimates were within the range of average professional learning intervention effects identified in the literature, they fell below our initial goal, which was based on the assumption that the effect of the booster intervention would likely be small compared to conventional professional learning effects.

A.3. About the professional learning program

All teachers participating in the research study (5 to 10 teachers per school from up to 20 schools[20]) were invited to complete AISWA's HIPs program. Upon completing the HIPs program, some schools were randomly assigned to receive a professional learning booster in the form of 2 AISWA-facilitated online webinars.

A.3.1. The High Impact Practices program

The HIPs program is an evidence-informed professional learning program that focuses on high-impact practices that have been shown to improve teaching and student outcomes. The high impact practices and strategies taught through the HIPs program align with those recommended by AERO, particularly in terms of explicit instruction and formative assessment. For more information, see the HIPs program overview and Section 2.1.

19 Note that due to the lack of evidence examining the effect of booster interventions following conventional professional learning on teacher-level outcomes, these reflect estimated effects for professional learning interventions in general, and for online professional learning interventions specifically (in line with the delivery method for the booster).

20 Recruitment closed with a total of 107 teachers and 17 leaders across 17 schools.

The HIPs program comprises:

- completing required pre-course readings
- completing required pre-course online learning
- attending 3 face-to-face workshops covering the high impact strategies
- completing post-workshop activities individually and as a group
- participating in at least one video recorded or live classroom observation (maximum of 2), with coaching and feedback by an AISWA HIPs facilitator.

The HIPs program incorporates a range of effective mechanisms of professional learning identified by the EEF, and would qualify as a balanced design (Sims et al., 2021). Examples include:

- managing cognitive load (use of examples and dual-coding)
- revisiting prior learning (review of prior learning at the start of each workshop and end of each module)
- setting goals (providing goal-setting templates and specifying goals for observation sessions)
- presenting credible information (suggestions supported by research, with links to additional resources)
- instruction on techniques (guided instruction for each technique)
- practical social support (pair-share and group discussion during workshops, prompts to engage as a group outside of workshops)
- modelling techniques (video examples)
- providing feedback (2 coaching and in-class observation and feedback sessions with an AISWA facilitator, with written feedback)
- providing prompts and cues (provided workbook and slides include prompts for teachers inside and outside workshop activities)
- encouraging self-monitoring (reflective journaling tasks in the workbook, prompts to reflect on implementation of new practices)
- prompting context-specific repetition (individual planning to use techniques in their lessons).

A.3.2. The booster intervention

The booster intervention was co-designed by AISWA and AERO, and comprised two 90-minute online professional learning sessions delivered by an AISWA facilitator to each school in the Booster group.

The booster design was informed by an initial literature review (see Section 1.4 and the Elements of successful boosters), followed by an iterative workshopping phase over several months of collaboration between AERO and AISWA. Once an agreed-upon structure was established, the content was populated by AISWA to align with the HIPs program, then reviewed and refined by both teams until it was ready for delivery by AISWA.

Teachers participated in the 2 booster intervention workshops with the same group of 5 to 10 teachers from their school. Incorporating effective mechanisms of professional learning, with a focus on revisiting prior learning, the purpose of the professional learning booster was to support teachers in embedding and sustaining evidence-based teaching practices over time.

The booster sessions provided an opportunity to:

- consolidate and revisit prior learning
- collaborate, discuss and problem-solve implementation challenges
- celebrate successes
- plan, set goals and discuss next steps.

In line with prior research (see the Booster intervention overview), the intervention was delivered to multiple teaching staff to facilitate collaboration, was not limited to a single session, and reinforced content and skills that were directly aligned with the initial professional learning program. As much as was possible within a brief intervention, the workshops were also designed to be responsive to teachers' needs and address challenges they may be experiencing in implementing evidence-based teaching practices. The booster intervention was also designed to incorporate effective mechanisms of professional learning identified by the EEF, and to qualify as a balanced design (Sims et al., 2021). Examples include:

- managing cognitive load (use of examples, limiting information presented simultaneously on slides)
- revisiting prior learning (review of prior learning from the HIPs program)
- setting goals (clear learning intentions for webinar)
- presenting credible information (suggestions supported by research, with links to additional resources)
- instruction on techniques (guided instruction for each technique)
- practical social support (small-group practice in breakout groups during webinar, suggestions for additional peer-support activities)
- modelling techniques (video examples)
- encouraging self-monitoring (reflection questions focused on in-context implementation experience)
- prompting context-specific repetition (individual action plan to use techniques in their lessons, with plans reviewed in small groups to help solve context-specific problems).

A.4. Data sources

Over the course of the study, data was collated from a total of 267 surveys and 45 interviews with teachers, leaders and AISWA HIPs facilitators (see Table 1). Primary outcome data was collected through self-report responses to teacher surveys, with process evaluation data collected across all sources using a mixture of open-ended questions, semi-structured interviews and additional data from fidelity checks.

Table 1: Data overview

Method	Data source	Booster group	Non-Booster group	Full sample
Surveys	**Teachers**	80	144	224
	Leaders	13	30	43
Interviews	**Teachers**	13	7	20
	Leaders	6	10	16
	Facilitators	N/A	N/A	9

Note: Data shown here are collapsed across timepoints.

A.4.1. Teacher surveys

All participating teachers completed an online self-report survey at the beginning of the study prior to delivery of the HIPs program (Baseline/T1), following the completion of the HIPs program (T2), and at the conclusion of the study, 6 months after the booster program (T3; see Table 2). These surveys were used to monitor changes in our outcomes of interest and to collect demographic and contextual information about each teacher, their school and their program experience.

Table 2: Teacher survey responses

Group	Timepoint			Total
	T1	T2	T3	
Booster	33	24	23	80
Non-Booster	62	35	47	144
Full sample	95	59	70	224

Note: A total of 107 teachers consented to participate in the study. T1: prior to the delivery of the HIPs program; T2: after the completion of the HIPs program; T3: after the completion of the booster.

Teacher demographics: Basic data about each respondent (school, name, email) was obtained to link responses across timepoints and track group allocation, with all identifying information replaced with anonymous participant and school IDs prior to analysis. At T1, teachers also provided information about their teaching experience, current role, and the year levels and subjects they taught.

Quality of use of evidence-based practices: Quality of use of evidence-based teaching practices was measured at each timepoint using AERO's explicit instruction and formative assessment rubrics. Together, these form a 20-item self-report rubric, with items scored either 0 to 2 or 0 to 3, depending on the number of differentiable levels in quality of practice use, for a total quality of practice score out of 49. This measure has been externally validated by the Assessment and Evaluation Research Centre (AERC), and was found to be sufficiently accurate to assess aggregated changes in teaching practice as a dependent variable (Robertson et al., 2024; see Appendix C for details). This was one of the primary outcome measures used in the study.

This measure assesses how teachers:

1. Identify how student knowledge and skills build in the learning area
2. Chunk learning for students
3. Use guided practice
4. Provide scaffolding to students
5. Use scaffolds to support student learning
6. Provide opportunities for independent practice of recently learned skills
7. Interact with students during practice
8. Make links between the skills and knowledge that has been explicitly taught
9. Set learning intentions
10. Set success criteria
11. Communicate learning intentions and success criteria
12. Design opportunities to collect evidence of student learning
13. Interpret evidence of student learning to inform decision-making
14. Review the suitability of evidence to inform a teaching decision
15. Make teaching decisions using evidence of student learning
16. Monitor student learning
17. Provide feedback
18. Make connections between feedback and learning intentions
19. Check for student understanding of feedback
20. Establish classroom climate in which students feel supported to learn from feedback.

Confidence to use evidence-based practices: Confidence to use evidence-based teaching practices was measured at each timepoint using 6 bespoke survey items with a 5-point Likert scale (Not at all confident to Very confident). These items were derived from a subset of the rubric items. This was one of the primary outcome measures used in the study.

This measure assesses how confident teachers are to:

1. Chunk learning for students
2. Set learning objectives and success criteria
3. Use guided practice
4. Provide opportunities for independent practice
5. Check for student understanding
6. Provide feedback.

Perceived student outcomes: Perceived student outcomes were measured at each timepoint using 3 bespoke survey items with a 5-point Likert scale (None to All). This was one of the primary outcome measures used in the study.

This measure assesses how many students in a teacher's current class:

1. Show active participation in their learning
2. Display confidence in their learning
3. Are performing at the expected standard of achievement for their year level.

Frequency of use of evidence-based practices: Frequency of use of evidence-based practices was measured at each timepoint using 20 bespoke survey items with a 5-point Likert scale (Never to Every lesson).

School culture: At T1 only, teachers were asked about their school culture around evidence use and readiness for change. School culture around evidence use was measured using 7 bespoke survey items with a 5-point Likert scale (Strongly disagree to Strongly agree) for a series of statements covering school policy and program decisions, teaching practice, support, and collaboration. Change readiness was measured using 3 bespoke survey items asking about teachers' desire and ability to make changes, as well as that of their colleagues.

HIPs program and booster intervention experience: At T2, after completing the HIPs program, teachers were asked to provide their perceptions of the HIPs program, and detail any enablers, barriers or issues that they experienced in program participation or implementation. Perceptions of the HIPs program were assessed using 5 bespoke survey items with a 5-point Likert scale (Strongly disagree to Strongly agree), for a series of statements covering the relevance, suitability, usability and overall program experience. Information about enablers and barriers in program participation and implementation was collected through 4 open-ended questions for thematic analysis. At T3, teachers in the Booster group were asked for corresponding information about their experiences with the Booster intervention, using the same set of measures.

Control variables: Across the survey, control and fidelity data were collected, primarily for use in assessing group balance, and to act as potential covariates for primary outcome analyses. Teachers were asked about the recency of their past professional learning experiences, familiarity with AERO and AISWA measures and resources, and other participating schools, and their HIPs program and booster intervention completion and participation.

A.4.2. Leader surveys

All participating leaders completed an online self-report survey at the same 3 timepoints as teachers (T1 before HIPs, T2 after HIPs and T3 at the conclusion of the study; see Table 3).

Table 3: Leader survey responses

Group	Timepoint			Total
	T1	T2	T3	
Booster	5	5	3	13
Non-Booster	12	10	8	30
Full sample	17	15	11	43

Note: There was one leader for each participating school, meaning responses were obtained for the full sample of 17 at T1.
T1: prior to the delivery of the HIPs program; T2: after the completion of the HIPs program; T3: after the completion of the booster.

Leader demographics: Basic data about each respondent (school, name, email) was obtained to link responses across timepoints and track group allocation, with all identifying information replaced with anonymous participant and school IDs prior to analysis. At T1, leaders also provided information about their education experience, current role and their involvement in professional learning oversight and classroom observation and teaching.

School-level changes: At T2 and T3, leaders were asked about any school-level changes to practice as a result of program participation. Data was collected using 3 open-ended response questions, probing implemented and planned changes to use of evidence-based practices.

School culture: At T1 only, leaders were asked about their school culture around evidence use and readiness for change. School culture around evidence use was measured using 5 bespoke survey items with a 5-point Likert scale (Strongly disagree to Strongly agree), for a series of statements covering school policy and program decisions, teaching practice, support, and collaboration. Change readiness was measured using 4 bespoke survey items asking about their desire and ability to make changes to practice across their school, and the desire and ability of their staff to make changes to their practice.

HIPs program and booster intervention experience: Although they did not directly participate in the HIPs program, at T2 (after its delivery), leaders were asked to provide their perceptions of the HIPs program, and detail any enablers, barriers or issues that they or their staff experienced in program participation or implementation. Perceptions of the HIPs program were assessed using 5 bespoke survey items with a 5-point Likert scale (Strongly disagree to Strongly agree), for a series of statements covering the relevance, suitability, usability and overall program experience. Information about enablers and barriers in program participation and implementation was collected through 4 open-ended questions for thematic analysis. At T3, leaders in the Booster group were asked for corresponding information about their experiences with the Booster intervention, using the same set of measures.

A.4.3. Interviews

Teachers, school leaders and facilitators participated in semi-structured interviews after the completion of the HIPs program in 2023, and again at the completion of the study in 2024 (T2 and T3, respectively) (see Table 4). Interviews were conducted in-person and online via Microsoft Teams depending on the preference of the school and ranged from 13 to 26 minutes in length. Interviews were recorded and audio was professionally transcribed using SmartDocs Transcription Services, with the quality checked by the research team. The purpose of these interviews was to collect rich qualitative data on the implementation and experience of the program, to identify enablers and barriers to practice change, and to assess which elements of the program were perceived as the most effective. The interview framework was based on the research questions, and at the end of each interview, participants were given the opportunity to provide any other feedback or comments that had not already been captured. Each of the AISWA HIPs facilitators was also interviewed at T2 to collect information about program delivery, perceived reception by participating staff, potential program changes and the components that were perceived as most or least effective. An equivalent interview was conducted at T3 with the facilitator who delivered each of the booster sessions, collecting corresponding information about booster delivery and reception.

Table 4: Interview responses

Group	T2 responses: Oct to Nov 2023	T3 reponses: Aug to Oct 2024	Total reponses
Booster	1 Leader 5 Teachers	5 Leaders 8 Teachers	6 Leaders 13 Teachers
Non-Booster	4 Leaders 3 Teachers	6 Leaders 4 Teachers	10 Leaders 7 Teachers
AISWA	8 Facilitators	1 Facilitator	9 Facilitators
Full sample	21	24	45

Note: Some teachers and leaders participated in interviews at both time points. The one facilitator interviewed at T3 was also among the 8 interviewed at T2 or similar. Some participants completed interviews at both timepoints. T2: after the completion of the HIPs program; T3: after the completion of the booster.

A.4.4. Fidelity data

After each session of the HIPs program and the booster intervention, facilitators collected fidelity data for each school (see Section 2.3). This data captured participant attendance and absences, the modules covered, time adequacy for each program activity, pre-session conversations, participant engagement, any delivery issues and additional feedback from facilitators or teachers attending the session. Attendance and engagement were good overall and were comparable across schools in the Booster group (see Table 5 for aggregated scores, and Table 6 for time for each component).

Table 5: Booster fidelity data

Booster session	Attendance	Engagement (/5)
Session 1	76.6%	4.6
Session 2	70.0%	4.4

Table 6: Booster fidelity data

During the webinar(s), did you have an opportunity to do the following?	Yes	Yes, but not as much as I would have liked to	No
Revisit some of the knowledge and skills you learned during the HIPs program	93.75%	6.25%	0%
Reflect on your progress in using high-impact practices in your teaching	81.25%	12.50%	6.25%
Reflect on your goals to use high-impact practices in your teaching	62.50%	25%	12.50%
Plan actions to use high-impact practices in your teaching	50%	37.50%	12.50%
Share learnings with colleagues about how you have been using high-impact practices in your teaching	62.50%	37.50%	0%
Troubleshoot problems in using high-impact practices in your teaching	62.50%	37.50%	0%

Note: All scores presented as percentages of the 16 Booster group teachers who provided responses to this survey section.

A.5. Sample description

A.5.1. Descriptive data

This study included 107 teachers and 17 leaders across 17 participating schools in Western Australia (see Section 2.1.3 for more information about participating schools).

Participant demographics: Of the 107 participating teachers and 17 participating leaders, baseline demographic data was obtained for 96 teachers and all leaders. Twenty-nine of the 96 teachers had leadership roles in addition to their teaching responsibilities (note that this was a different individual to the 'leader' participant at each school), and 31 taught exclusively primary school classes, 63 taught secondary and 2 taught classes across both levels. On average, teachers in the sample had been working in education for a total of 12.5 years and been at their current school for roughly half of that time. Leaders had an average of 25.5 years of experience and had been at their current school for roughly 9 years on average (see Table 7). At the beginning of the project teachers were asked if they had previously participated in any professional learning on instructional teaching practices, with 29% indicating they had in the past year, 17% indicating they had more than one year ago and 54% indicating that they had never participated in professional learning of that sort.

Table 7: Participant years of experience

Years of experience	Teachers (years)	Leaders (years)
In current role	4.07	6.41
In current school	6.28	8.71
In education field	12.50	25.50

School culture: At T1, all teachers and leaders in the study provided ratings of their school culture around evidence use and their readiness for change. Responses were mostly positive about overall school culture around evidence use, with leaders providing slightly stronger endorsement on average, but also showing greater agreement with the idea that their school would benefit from changing instructional practice (see Table 8).

Table 8: Participant ratings of school culture

Items (Strongly disagree to Strongly agree)	Teachers (/5)	Leaders (/5)
At my school, we use evidence when deciding on school policies and programs.	3.71	4.06
At my school, leaders share and discuss evidence that could improve our practices.	3.67	4.00
At my school, leaders encourage me to use evidence to determine whether a practice works.	3.54	N/A
At my school, support is available to help teachers change their practice.	3.69	4.18
At my school, colleagues and I discuss evidence that could improve our practices.	3.87	3.88
My colleagues encourage me to use evidence to change my practice.	3.19	N/A
My school could benefit from changing instructional practices used with students.	3.84	4.29

Note: Respondents indicated their level of agreement on a 5-point scale, with 1 = Strongly disagree and 5 = Strongly agree. A higher value indicates a higher level of agreement.

Readiness for change responses indicated that most teachers and leaders wanted and felt able to make changes to their (or their school's) teaching practice, although a proportion of those indicated that they would benefit from some help. Teachers were less confident in their colleagues' ability to make changes. However, leaders reported that their staff predominantly wanted to make changes and were ready to do so, although most would benefit from help (see Table 9 and Table 10 for teacher and leader responses, respectively).

Table 9: Teacher-reported readiness for change

Teachers (*n* = 94)	Not sure	No	Yes
Do you want to make changes to your practice?	6%	1%	93%
Are you ready and able to change your practice?	9%	0%	52% (Yes) 39% (Yes, but with help)
Do you think most teachers at your school are ready and able to make changes to their practice?	39%	23%	38%

Table 10: Leader-reported readiness for change

Leaders (*n* = 17)	Not sure	No	Yes
Do you want to make changes to teaching practices in your school?	6%	0%	94%
Are you ready and able to support teachers at your school to make changes to practice?	6%	0%	76% (Yes) 18% (Yes, but with help)
Do teachers at your school want to make changes to their practice?	18%	12%	70%
Are teachers at your school ready and able to change their practice?	0%	0%	30% (Yes) 70% (Yes, but with help)

Program experience: Teachers and leaders were generally very positive about their experience of the HIPs program (see Table 11), giving high ratings across items. Teachers showed the greatest endorsement of the program's relevance, and the lowest endorsement of its ease of use, although all ratings were very high on average, falling between Agree and Strongly agree on a 5-point Likert scale. Teachers and leaders were also very positive about the booster intervention, although mean ratings were marginally lower on average (albeit still very high overall), and only 2 of the 5 leaders from booster schools provided data for these questions (see Table 12).

Table 11: HIPs program experience

The HIPs program I received...	Teachers (/5)	Leaders (/5)
Seems relevant	4.72	4.69
Seems suitable	4.69	4.77
Seems doable	4.50	4.77
Seems easy to use	4.33	4.54
Is something I would recommend	4.52	4.77

Note: Respondents indicated their level of agreement on a 5-point scale, with 1 = Strongly disagree and 5 = Strongly agree. A higher value indicates a higher level of agreement.

Table 12: Booster intervention experience

The professional learning booster I received...	Teachers (/5)	Leaders (/5)
Seems relevant	4.31	5.00
Seems suitable	4.44	5.00
Seems doable	4.44	5.00
Seems easy to use	4.38	5.00
Is something I would recommend	4.31	4.50

Note: Respondents indicated their level of agreement on a 5-point scale, with 1 = Strongly disagree and 5 = Strongly agree. A higher value indicates a higher level of agreement.

School-level changes: At T2 and T3, leaders reported on school-level changes to practice as a result of program participation. At T2, leaders at two-thirds of participating schools (9/13) reported that they had not yet made any changes but were planning to do so in the near future, and leaders at 4 out of 13 schools reported that they had made school-level changes to support the use of evidence-based practices. By T3, half of the leaders who answered the survey (5/10) reported they were still planning to make changes but were yet to do so, 4 out of 10 reported that they had made school-level changes, and one school leader indicated that no changes were made or planned.

A.5.2. Group balance checks

After randomisation, baseline group balance was compared between the Booster and Non-Booster groups on each of the following variables:

- school ICSEA
- school size (student enrolments)
- teacher years of experience (in role, at school, in total)
- school culture
- change readiness
- confidence to use evidence-based practices
- quality of use of evidence-based practices
- perceived student outcomes
- frequency of use of evidence-based practices
- pattern and extent of missing data.

No differences in baseline scores were found between the Booster and Non-Booster group for any tested variables (all $p > 0.05$). This check was repeated, using schools that declined the booster as a separate group (see Appendix A.2.1), which also indicated that there were no significant baseline differences on any variable (all $p > 0.05$). These baseline group balance checks were conducted using *t*-tests, Little's MCAR test, and Chi-Square goodness-of-fit tests, as appropriate for each data type. As an additional check, we ran ITT analyses, using the original group allocation for each school, regardless of whether they participated in the booster. These additional analyses revealed the same overall pattern of results, suggesting that booster-refusal had not unduly influenced our findings (see Appendix B.3).

A.6. Design

A.6.1. Statistical analysis

We used multilevel linear mixed models to determine whether booster participation improved our outcomes of interest more than the professional learning alone. These models adjusted for clustering, the tendency for individuals within a 'cluster' to be more similar to each other than members of a different cluster. Clusters in this study were the individual schools participating in the study.

These models also controlled for other variables or 'covariates' that are plausibly related to our outcome of interest. We used a stepwise model-building procedure for each dependent variable to avoid overfitting, based on the 'purposeful selection' approach described by Hosmer et al. (2013). All potential covariates were tested for univariate associations with the dependent variable. All covariates that had a univariate association with the outcome variable of $p < 0.25$ were included in an initial model, which was iterated on repeatedly, with the least significant predictor removed at each step until a final model was reached, containing timepoint, group and any significant covariates. We tested the fit of this model, performing a series of model simplification checks to determine whether the inclusion of covariates and random effects was appropriate. Finally, we ran a series of planned contrasts on the final model to answer our research questions.

The covariates tested were the same for each dependent variable, with values drawn from each teacher's survey responses for teacher-level covariates, and from leader surveys and publicly available data for school-level covariates. The covariates we tested for each model were as follows:

- Teacher-level:
  - role
  - year levels taught (primary/secondary/both)
  - prior professional learning
  - experience (role/school/total teaching experience)
  - change readiness
  - perceived school culture around evidence use
  - baseline confidence to use evidence-based practices
  - baseline perceived student outcomes
  - baseline quality of use of evidence-based practices
  - HIPs program experience
  - post-HIPs program collaboration (presence/frequency/utility)
  - booster experience (booster group analyses only)
  - booster components (booster group analyses only).
- School-level:
  - school size
  - ICSEA
  - year levels (primary/secondary/combined).

Covariates related to the booster were only included for a secondary set of subgroup analyses, using only Booster group teachers (and no grouping factor), to determine whether there were any meaningful contributions to outcomes within this intervention group (see Appendix B.1).

As an additional check, if more than 5% of measurements of a variable were missing, we considered both the extent and pattern of missing data to determine next steps. Based on Little's test (Little, 1988), if the data did not appear to be missing completely at random, we planned to use multiple imputation to estimate missing values and include analyses both with and without the imputed data. However, all missing data was determined to be missing at random, so no imputed data was used in any analyses.

A.6.2. Thematic analysis

Qualitative data was collated and analysed in NVivo (Version 14) and analysed using thematic analysis (Braun & Clarke, 2012). The research team familiarised themselves with the dataset by reading and re-reading the interview transcripts and taking notes about analytic ideas and insights. In the coding phase, segments of data were coded combining inductive and deductive approaches. Initial themes were identified and reviewed by the broader research team. Finally, the themes were refined and named, building a clearer story from the data. Multiple research team members worked on the analysis and development of the coding framework. To ensure the reliability of the results, a subset of interviews was coded by 2 research team members. From this process, inter-rater reliability was determined to be satisfactory ($\kappa = 0.73$). Rigour of the data was achieved through transparent and complete recording, with the data organised around themes, including direct quotes, where appropriate.

Appendix B: Additional results

B.1. Primary outcome analyses

Based on the recommendations of the EEF (2022) and the What Works Clearinghouse (2022), we used multilevel linear mixed modelling to test for the contribution of time and booster participation to teacher outcomes, with individual teachers and schools as random effects. These models adjusted for the clustering of teachers within schools, and for any covariates that were found to be significantly related to each outcome (see Appendix A.6.1 for the model-building procedure and list of tested covariates). For each outcome variable, we also ran an additional 'booster-only' model, including only teachers who participated in the booster intervention, with each individual's reports on the presence of booster components as additional potential covariates (see Appendix A.4.4), as well as a binary variable capturing whether the same facilitator was assigned to each school for the HIPs program and Booster program. These additional models were intended as a check for any meaningful interactions or important booster components that would be missed in the full model with Non-Booster group teachers that lacked data for these items. The results of each analysis are reported in this appendix, with interpretation and implications for major results in Section 3 and Section 4, respectively. Item-level results for the components of each scale are presented visually to illustrate trends, but individual components were not used as outcomes in analyses as they lack the precision to give meaningful results.

B.1.1. Quality of use of evidence-based practices

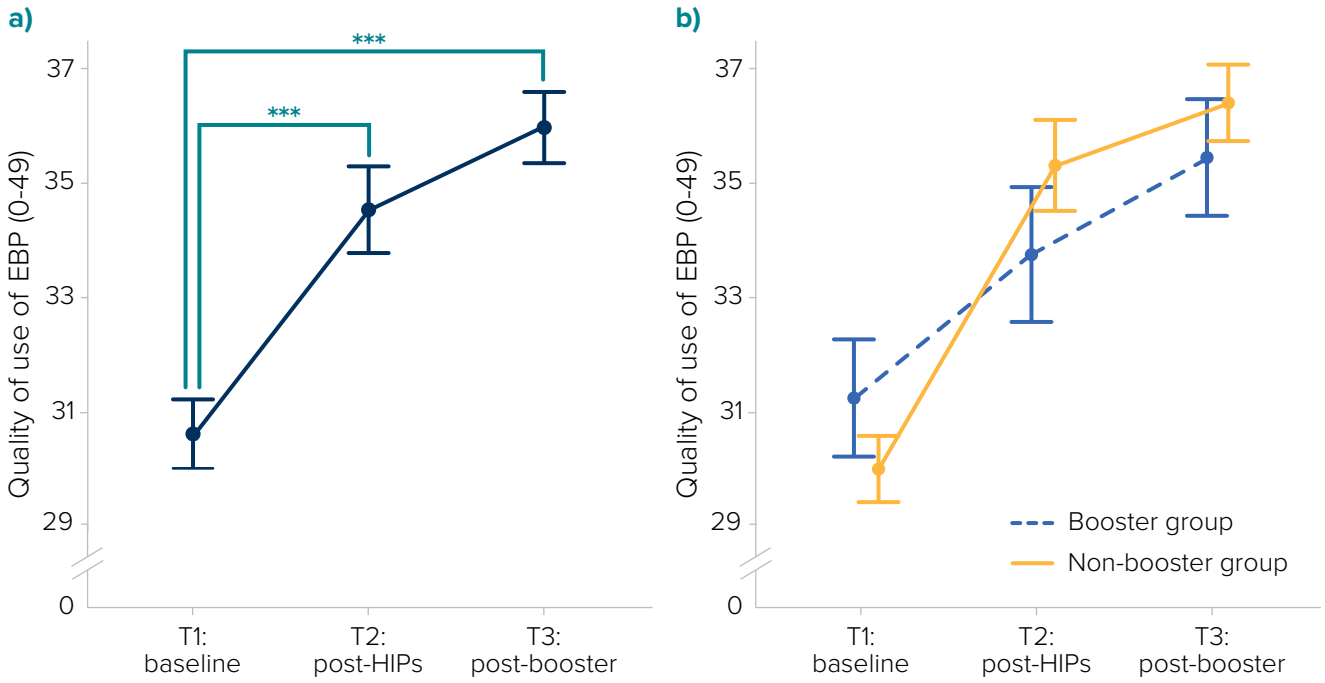
Quality of use of evidence-based practices, baseline quality of use scores and one school culture item ('My school could benefit from changing instructional practices used with students') were found to be significant covariates and were included in the final model (see Table 13 for final model results).

Table 13: Final model results (Quality of use of evidence-based practices)

Variable	Estimate	Significance
Time	1.31	$p < 0.001^{***}$
Group	0.47	$p > 0.05$
Baseline score	0.69	$p < 0.001^{***}$
School culture (My school could benefit from changing instructional practices used with students.)	1.13	$p = 0.008^{**}$

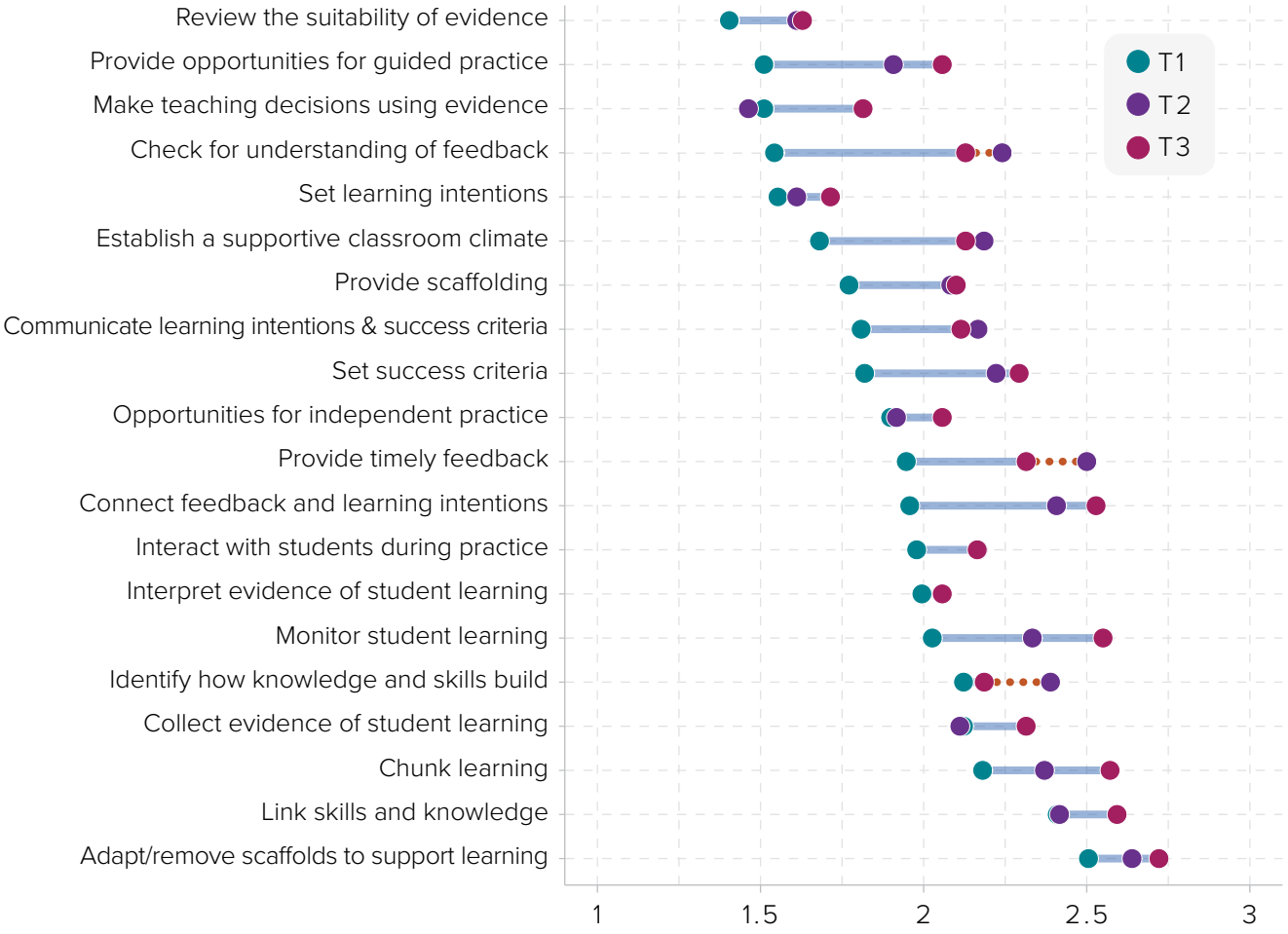
The analysis showed a significant overall effect of time, indicating that teachers' outcomes improved over the course of the study, but no effect of booster participation. The baseline quality of evidence-based practice use was positively associated with overall scores, indicating that higher baseline scores predicted higher scores across timepoints. The school culture covariate was positively associated with the outcome variable, indicating that teachers who more strongly endorsed the need for changing instructional practices were more likely to have higher-quality of practice scores. The planned contrast testing the interaction between groups and T2 to T3 changes was not significant. The additional booster-only analysis did not reveal any meaningful differences within the Booster group, and none of the booster-related covariates were found to be significantly associated with the outcome variable. The changes in quality of use of evidence-based practices and quality of use of evidence-based strategies over time across the two groups are presented in Figure 8 and Figure 9 respectively.

Figure 8: Quality of use of evidence-based practices over time



Note: Panel (a) shows quality of use of evidence-based practice across both groups combined, with panel (a) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line. Analysis based on 169 observations of 72 teachers. Error bars represent ± 1 standard error.

Figure 9: Quality of use of evidence-based strategies over time



Note: This shows rubric scores for individual strategies (components of evidence-based practice). Raw scores for individual rubric items vary from 0-2 or 0-3, but in this figure, all scores have been normed to a 0 to 3 range for comparison, with higher scores representing better practice. Raw scores were used for all analyses and other figures. Solid bars in blue represent an improvement compared to the previous timepoint, dotted lines in orange represent a decrease.

B.1.2. Confidence to use evidence-based practices

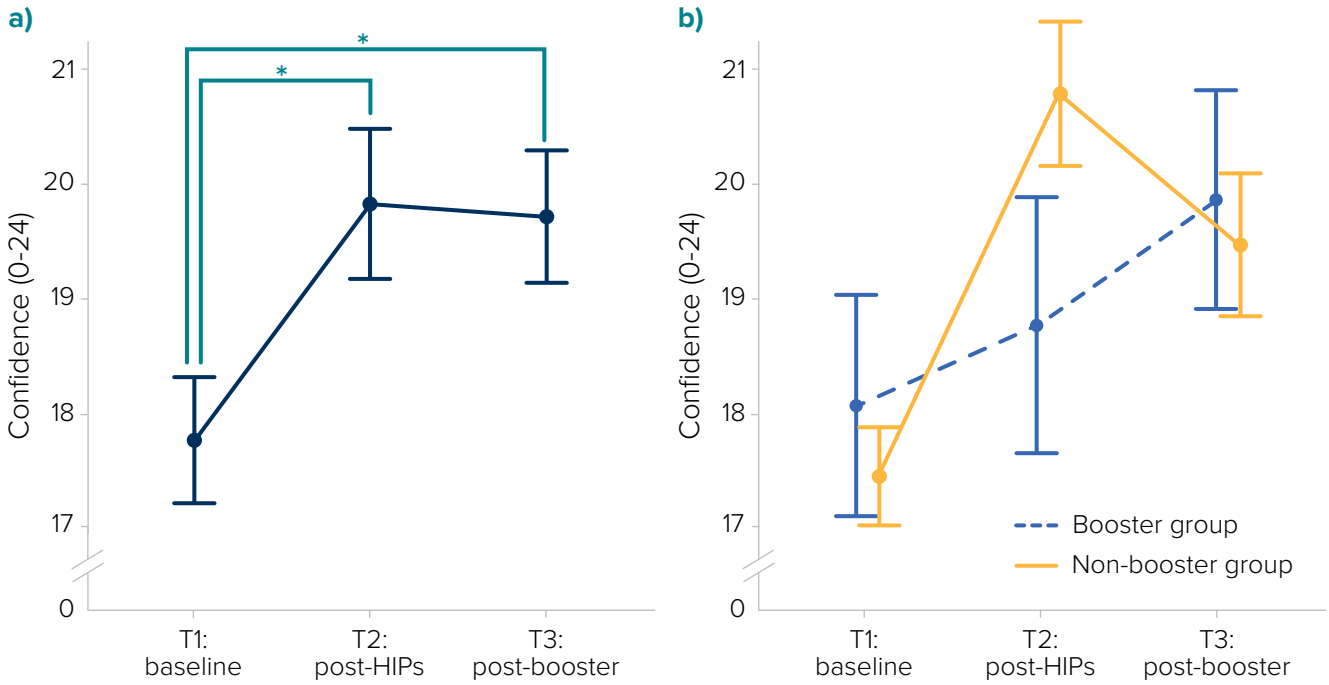
For confidence to use evidence-based practices, baseline scores and one teacher experience component (years spent in current role) were found to be significant covariates and were included in the final model (see Table 14 for final model results).

Table 14: Final model results (Confidence to use evidence-based practices)

Variable	Estimate	Significance
Time	0.71	$p = 0.020^{*}$
Group	0.33	$p > 0.05$
Baseline score	0.49	$p < 0.001^{***}$
Experience Q2 (How long have you been in your current main role at [your school]?)	-0.26	$p < 0.001^{***}$

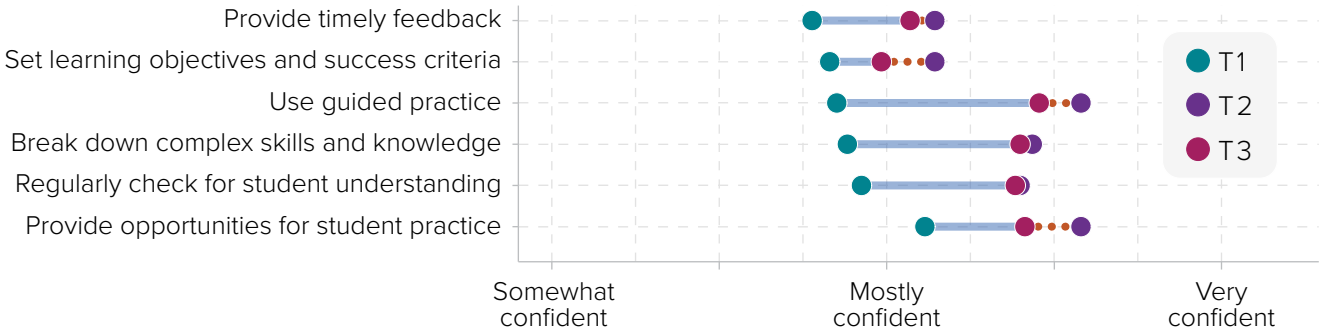
The analysis showed a significant overall effect of time, indicating that teacher outcomes improved over the course of the study, but no effect of booster participation. Baseline confidence in using evidence-based practices was positively associated with overall scores, meaning that higher baseline scores predicted higher overall scores. The experience covariate was negatively associated with the outcome variable, indicating that teachers who had been in their role for longer were more likely to have lower confidence scores. The planned contrast testing the interaction between group and T2 to T3 changes was not significant. The additional booster-only analysis did not reveal any meaningful differences within the Booster group, and none of the booster-related covariates were found to be significantly associated with the outcome variable. Changes in the levels of confidence to use evidence-based practices and evidence-based strategies over time across the 2 groups are mapped in Figure 10 and Figure 11 respectively.

Figure 10: Confidence to use evidence-based practices over time



Note: Panel (a) shows confidence to use evidence-based practice across both groups combined, with panel (b) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line. Analysis based on 185 observations of 80 teachers. Error bars represent ± 1 standard error.

Figure 11: Confidence to use evidence-based strategies over time



Note: This figure shows confidence ratings for individual strategies (components of evidence-based practice). Solid bars in blue represent an improvement compared to the previous timepoint, dotted lines in orange represent a decrease.

B.1.3. Perceived student outcomes

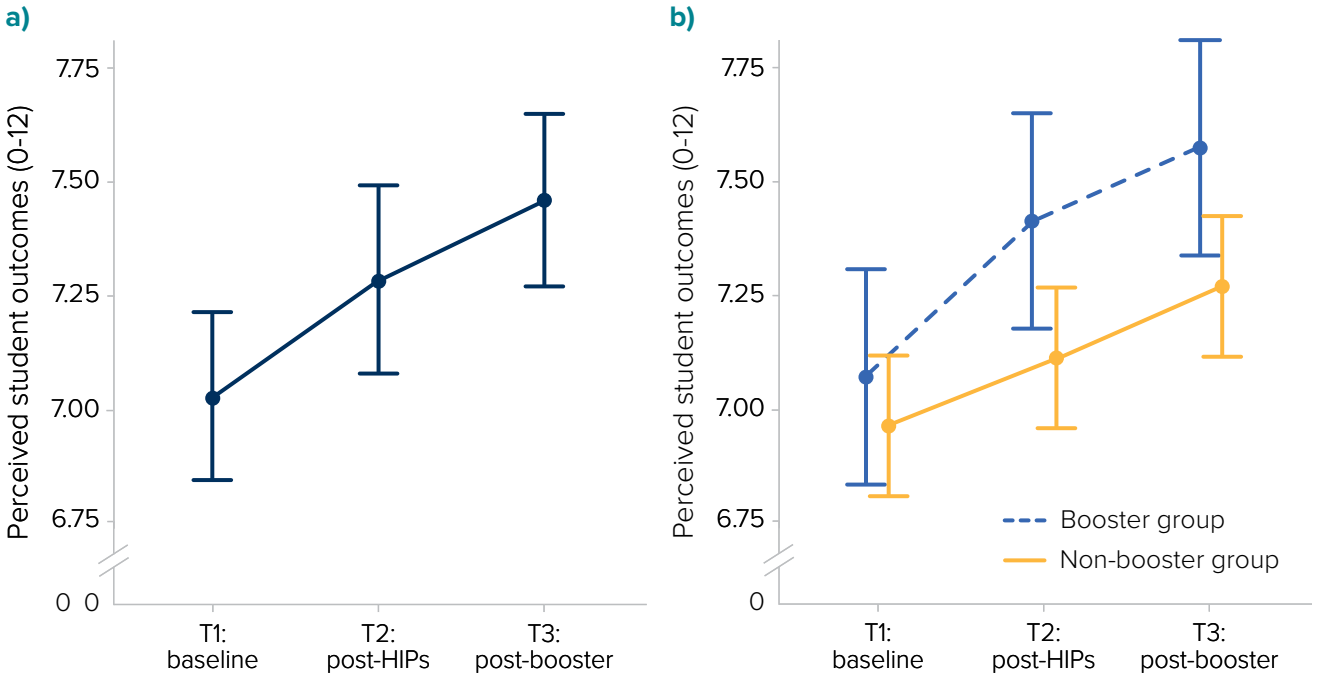
For perceived student outcomes, baseline scores, school culture items related to evidence-use and one specific school culture item ('My school could benefit from changing instructional practices used with students') were found to be significant covariates and were included in the final model (see Table 15 for final model results).

Table 15: Final model results (Perceived student outcomes)

Variable	Estimate	Significance
Time	0.14	$p > 0.05$
Group	0.28	$p > 0.05$
Baseline score	0.66	$p < 0.001$***
School culture items assessing evidence use	0.07	$p = 0.049$*
School culture (My school could benefit from changing instructional practices used with students)	-0.21	$p = 0.036$*

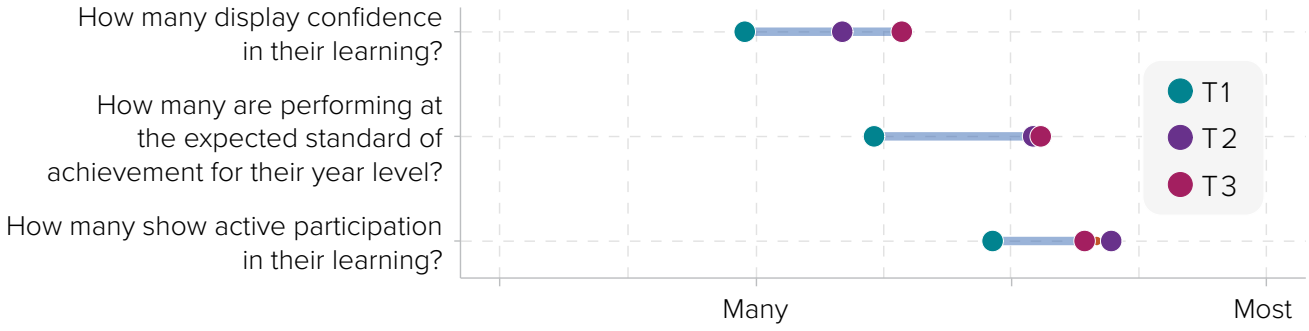
The analysis showed no significant effect of time or booster participation. Baseline perceived student outcomes were positively associated with overall scores, meaning that higher baseline scores predicted higher overall scores. School culture items that assessed evidence use were positively associated with overall scores, indicating that better use of evidence within a school was associated with higher perceived student outcomes, albeit weakly. The additional school culture covariate was negatively associated with the outcome variable, indicating that teachers who more strongly endorsed the need for changing instructional practices were likely to have lower ratings of perceived student outcomes. The planned contrast testing the interaction between group and T2 to T3 changes was not significant. The additional booster-only analysis did not reveal any meaningful differences within the Booster group, and none of the booster-related covariates were found to be significantly associated with the outcome variable. Figure 12 maps perceived student outcomes over time across the two groups.

Figure 12: Perceived student outcomes over time



Note: Panel (a) shows perceived student outcomes across both groups combined, with panel (b) showing outcomes for the Booster group in a dashed blue line, and the Non-Booster group in a solid yellow line. Analysis based on 154 observations of 66 teachers. Error bars represent ± 1 standard error.

Figure 13: Perceived student outcomes over time

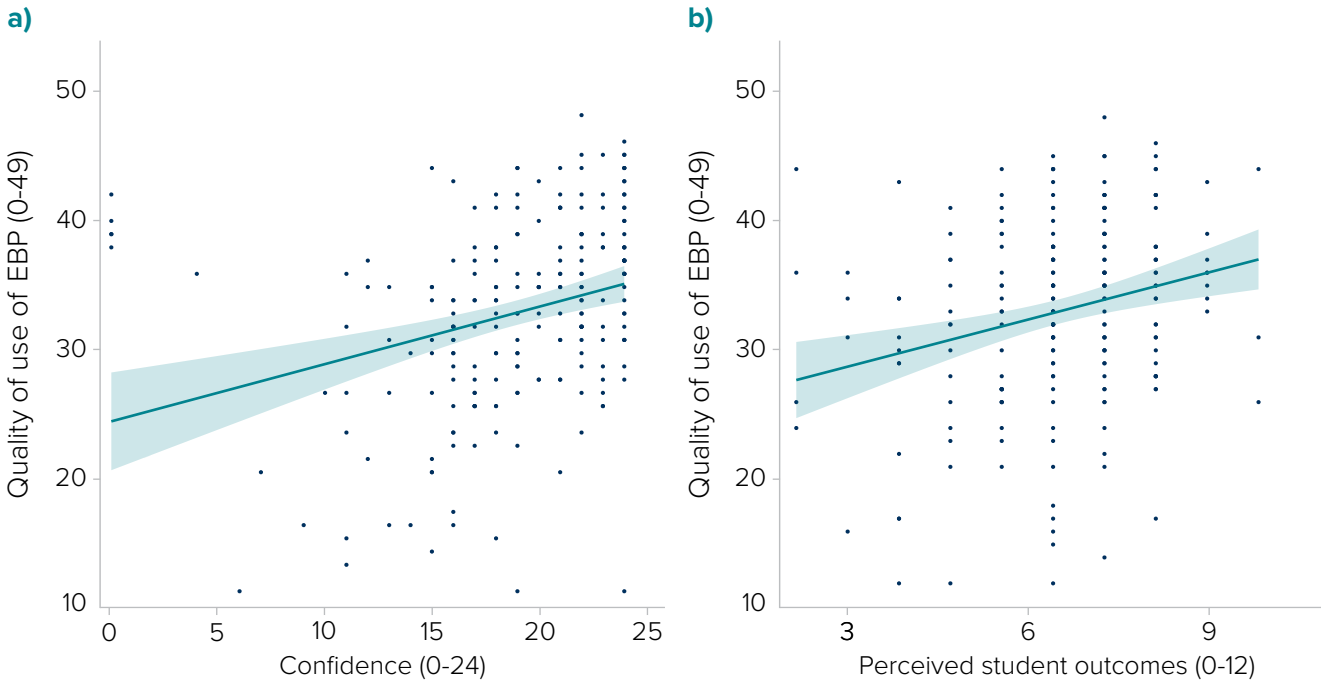


Note: Solid bars in blue represent an improvement compared to the previous timepoint, dotted lines in orange represent a decrease.

B.2. Inter-outcome associations

To determine the extent to which our primary outcomes were associated with one another, we calculated each of the 3 pairwise correlations between them. These indicated that quality of use of evidence-based practices was significantly associated with confidence to use evidence based-practices ($r = 0.29$, $p < 0.001^{***}$) and with perceived student outcomes ($r = 0.25$, $p < 0.001^{***}$), but that confidence and perceived student outcomes were not significantly associated with one another ($r = 0.09$, $p > 0.05$). Better quality of use of evidence-based practices was associated with increased confidence and better perceived student outcomes (see Figure 14).

Figure 14: Significant inter-outcome associations



Note: Panel (a) shows the association between quality of use of evidence-based practices and confidence to use evidence based-practices. Panel (b) shows the association between quality of use of evidence-based practices and perceived student outcomes. The shaded areas in blue represent the 95% confidence interval for each association.

B.3. Intention-to-treat analyses

To test whether schools that refused the booster differed from the rest of the booster sample and could have unduly biased our results, we compared their baseline characteristics with the rest of the sample (see Appendix A.2.1 and Appendix A.5.2) and conducted additional ITT analyses for each of our primary outcome variables. The process for these ITT analyses was identical to the primary outcome analyses, beginning with the same model-building procedure and tests for potential covariates. The only difference was that data from teachers in the schools that had refused the booster was recoded to the Booster group, in accordance with the initial random allocation. This meant that if our results significantly depended on the allocation of the booster-refused schools, then these ITT analyses could identify a different set of significant covariates and yield very different patterns of results to the primary analyses.

However, these ITT analyses yielded extremely similar results to the primary analyses, identifying an identical set of significant covariates for each primary outcome variable. The pattern of effects across groups and over time was also completely unchanged. As in the primary outcome analyses, there were no significant group differences, and no significant interactions between group and time, further suggesting that the booster intervention had not significantly affected teacher outcomes. See Table 16 for a comparison between the primary outcome analysis results and the ITT analysis results. Overall, the same pattern of results was obtained with the schools that refused the booster included in the Booster group, indicating that their refusal is unlikely to have influenced our findings.

Table 16: Comparison of primary analysis results and intention-to-treat analysis results

Primary outcome variable	Model variable	Primary outcome analyses		ITT analyses	
		Estimate	*p*-value	Estimate	*p*-value
Quality of use of evidence-based practices	Time	1.31	<0.001***	0.37	<0.001***
	Group	0.47	0.562	0.53	0.506
	Baseline score	0.69	<0.001***	0.71	<0.001***
	School culture Q7	1.13	0.008**	1.13	0.007**
Confidence to use evidence-based practices	Time	0.71	0.020*	0.36	0.001**
	Group	0.33	0.643	0.31	0.645
	Baseline score	0.49	<0.001***	0.48	<0.001***
	Experience in current role	-0.26	<0.001***	-0.25	0.002**
Perceived student outcomes	Time	0.14	0.179	0.17	0.145
	Group	0.28	0.301	0.08	0.176
	Baseline score	0.66	<0.001***	0.66	<0.001***
	School culture (evidence-use)	0.07	0.049*	0.06	0.048*
	School culture Q7	-0.21	0.036*	-0.20	0.044*

Note: Model variables include both fixed effects for group and time, as well as covariates identified as significant through the model-building procedure. ITT analyses were conducted in the same way as the primary outcome analyses, except that the group allocation variable was based on the initial random allocation, with schools that refused the booster included in the Booster group (see Appendix A.2.1).

Appendix C: AERO rubric validation

AERO developed 2 sets of rubrics to support assessment of teaching practices in the areas of explicit instruction (EI) and formative assessment (FA). The AERC at the University of Melbourne was commissioned by AERO to conduct research to refine, trial and validate these measures. The EI and FA Rubrics were aligned to the relevant AERO Tried and Tested practice guides (AERO, 2021a, 2021b, respectively) and were designed to measure the general use of these practices, which are not necessarily linked to any particular program or professional learning. The rubrics were intended to serve the following purposes:

1. provide feedback to individuals about their practice
2. summarise practice across a service, system or sector using aggregated data from multiple individuals
3. measure changes in an individual's practice over time as part of experimental study designs.

The rubrics were designed for use in both primary and secondary school settings across all education sectors in Australia (i.e., government, independent and Catholic school settings) by both teachers and others familiar with the teacher's practice, such as colleagues and school leaders. To ensure the tool was easy to use, the typical rubric structure was adapted so that indicative behaviours were grouped into sub-capabilities with consistent criteria within each sub-capability. Typically, the levels of quality within a rubric are different for each behaviour and describe levels of increasing competence. An additional option was added to allow respondents to indicate whether they did not have the opportunity to use this strategy or chose not to use it.

The rubric review by AERC included 5 stages:

1. Rubric review
2. Survey development
3. Field trial
4. Data analysis
5. Refinement.

Rubric review: For the review stage, there were 3 rounds of expert review. First, a concept mapping exercise in which the key components of each construct found in the literature were qualitatively mapped to the capabilities and indicators of the rubrics using expert judgement. This process helped ensure that there was adequate construct coverage of both sets of rubrics prior to the psychometric validation of the rubrics throughout the field trial. Second, a series of rubric review workshops were conducted with subject matter experts from the Faculty of Education at the University of Melbourne. The purpose of the workshops was to refine the sets of rubrics in light of their academic expert opinion about both constructs and rubric design. Third, feedback was sought from AERO stakeholders on the FA and EI rubrics in light of the amendments proposed from the concept mapping and workshops review activities.

Survey development: To empirically validate the modified FA and EI rubrics, an online validation survey was developed for each practice area, with multiple-choice versions of each rubric indicator, and teachers tasked with selecting the option that best described their current practice.

Field trials: Two field trials were conducted using modified versions of the rubrics developed during the rubric review step, for use in surveys. In the first trial, 104 teachers completed the EI rubric and 85 completed the FA rubric. In the second trial, 83 teachers who had completed both rubrics as part of another research trial consented to the use of their data in the validation.

Data analysis: The sensitivity, sequencing and construct validity of each rubric component were tested, as were the reliability and convergent validity of the rubric overall. Analyses were predominantly unidimensional Rasch analyses using Partial Credit Models, to examine the general psychometric properties of the rubrics, including item fit statistics (mainly the mean squared statistics), reliability, overall item difficulties, step difficulties, item discrimination and Classical Test Theory statistics.

Refinement: At the conclusion of the validation process, AERC provided recommendations to adjust the rubrics and their survey equivalents, based on their analysis. Adjustments aimed to improve the clarity of items and concepts, as well as increase the validity of the instruments.

Overall, the validation work indicated that the rubrics are suitable for use for monitoring practice at an aggregated level. This includes use as pre- and post-measures within an experimental design, provided significance testing was restricted to the aggregated scores. In their current form, the rubrics can be used to assess whether or not practice has changed, but should not be used to quantify these changes (an increase of 10 would not indicate twice as much improvement as an increase of 5; both increases would just indicate that practice has improved). With reliabilities between 0.75 and 0.82, the EI and FA rubrics were found to be sufficiently robust to be used as individual instruments, or as a combined instrument. Reliability was highest for the combined instrument (0.85), likely due to the greater number of items. The EI and FA rubrics were found to measure the intended constructs, although the analysis was unable to establish the extent to which the self-report ratings reflected teachers' practice.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). *Rethinking the use of tests: A meta-analysis of practice testing. Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Anshu, A. H. (2019). Impacts of action research-oriented upgrading training on initial and early reading comprehension performances of students. *International Journal of Education and Literacy Studies, 7*(1), 31–41. https://doi.org/10.7575/aiac.ijels.v.7n.1p.31

Australian Education Research Organisation. (2021a). *Formative assessment: Know where your students are in their learning*. https://www.edresearch.edu.au/guides-resources/practice-guides/formative-assessment-practice-guide-full-publication

Australian Education Research Organisation. (2021b). *Explicit Instruction: Know how to teach your students*. https://www.edresearch.edu.au/guides-resources/practice-guides/explicit-instruction-practice-guide-full-publication

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol. 2, pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Chipchase, L. S., Cavaleri, R., & Jull, G. (2016). Can a professional development workshop with follow-up alter practitioner behaviour and outcomes for neck pain patients? A randomised controlled trial. *Manual Therapy*, *25*, 87–93. https://doi.org/10.1016/j.math.2016.06.010

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Crowther, D. T., & Cannon, J. R. (2002). Professional development models: A comparison of duration and effect. *Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science*, 1–16.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

Dewing, J., Mathews, C., Cloete, A., Schaay, N., Simbayi, L., & Louw, J. (2014). Lay counselors' ability to deliver counseling for behavior change. *Journal of Consulting and Clinical Psychology, 82*(1), 19–29. https://doi.org/10.1037/a0034659

Dong, N., Reinke, W. M., Herman, K. C., Bradshaw, C. P., & Murray, D. W. (2016). Meaningful effect sizes, intraclass correlations, and proportions of variance explained by covariates for planning two-and three-level cluster randomized trials of social and behavioral outcomes. *Evaluation Review, 40*(4), 334–377. https://doi.org/10.1177/0193841X16671283

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0

Education Endowment Foundation. (2022). *Statistical analysis guidance for EEF evaluations*.

Edwards Groves, C., & Rönnerman, K. (2013). Generating leading practices through professional learning. *Professional Development in Education, 39*(1), 122–140. https://doi.org/10.1080/19415257.2012.724439

Elias, M. J. (2010). *Sustainability of social-emotional learning and related programs: Lessons from a field study*. https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.

Frank, B., & Kluge, A. (2018). Complex cognitive skill retention: The roles of general mental ability and refresher interventions in a simulated vocational setting. *Journal of Computer Assisted Learning, 34*(5), 471–481. https://doi.org/10.1111/jcal.12251

Fullan, M. (2006). The future of educational change: System thinkers in action. *Journal of Educational Change, 7*(3), 113–122. https://doi.org/10.1007/s10833-006-9003-9

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Gesel, S. A., LeJeune, L. M., Chow, J. C., Sinclair, A. C., & Lemons, C. J. (2021). A meta-analysis of the impact of professional development on teachers' knowledge, skill, and self-efficacy in data-based decision-making. *Journal of Learning disabilities, 54(*4), 269–283. https://doi.org/10.1177/0022219420970196

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7–8), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher, 42*(9), 476–487. https://doi.org/10.3102/0013189X13512674

Hill, H. C., & Papay, J. P. (2022). *Building better PL: How to strengthen teacher learning*. The Research Partnership for Professional Learning. https://annenberg.brown.edu/rppl/what-works

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Lord, P., Rabiasz, A., Roy, P., Harland, J., Styles, B., & Fowler, K. (2019a). *Evidence-based literacy support: The 'Literacy Octopus' trial: Evaluation report and executive summary with addendum*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-literacy-octopus-communicating-and-engaging-with-research

Lord, P., Rabiasz, A., & Styles, B. (2019b). '*Literacy Octopus' Dissemination Trial: Evaluation Report and Executive Summary with addendum*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-literacy-octopus-communicating-and-engaging-with-research

Miller, I., Crosland, K. A., & Clark, H. B. (2014). Behavioral skills training with teachers: Booster training for improved maintenance. *Child & Family Behavior Therapy, 36*(1), 19–32. https://doi.org/10.1080/07317107.2014.878176

Moore, D., Proctor, R., Benham-Clarke, S., Gains, H., Melendez-Torres, G., Axford, N., Rogers, M., Anderson, R., Hall, D., Hawkins, J., Berry, V., Forbes, C., & Lloyd, J. (2024). *Review of Evidence on Implementation in Education*. https://pearl.plymouth.ac.uk/pms-research/1128

Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Jaeger, A. (2019). Retrieval practice in classroom settings: A review of applied research. *Frontiers in Education, 4*(5), 1–16. https://doi.org/10.3389/feduc.2019.00005

Morina, F., Fütterer, T., Hübner, N., Zitzmann, S., & Fischer, C. (2025). Effects of online teacher professional development on teacher-, classroom-, and student-level outcomes: A meta-analysis. *Computers & Education, 228*, Article 105247. https://doi.org/10.1016/j.compedu.2025.105247

Owen, K. L., Hunter, S. H., Watkins, R. C., Payne, J. S., Bailey, T., Gray, C., Hastings, R. P., & Hughes, J. C. (2021). Implementation support improves outcomes of a fluency-based mathematics strategy: A cluster-randomized controlled trial. *Journal of Research on Educational Effectiveness, 14*(3), 523–542. https://doi.org/10.1080/19345747.2021.1875526

Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in psychology, 5*(286), 1–5. https://doi.org/10.3389/fpsyg.2014.00286

Prenger, R., Poortman, C. L., & Handelzalts, A. (2017). Factors influencing teachers' professional development in networked professional learning communities. *Teaching and Teacher Education, 68*, 77–90. https://doi.org/10.1016/j.tate.2017.08.014

Riley-Tillman, T. C., & Eckert, T. L. (2001). Generalization programming and school-based consultation: An examination of consultees' generalization of consultation-related skills. *Journal of Educational and Psychological Consultation, 12*(3), 217–241. https://doi.org/10.1207/S1532768XJEPC1203_03

Robertson, P., Sun, S. Z., Gillis, S., Hopfenbeck, T. N., & Rice, S. (2024). *Provision of robust measures: Explicit instruction and formative assessment rubrics* [Unpublished report prepared for the Australian Educational Research Organisation]. The University of Melbourne.

Rose, D. J., & Church, R. J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education, 8*, 5–35. https://doi.org/10.1023/A:1022860606825

Rusby, J. C., Jones, L. B., Crowley, R., Smolkowski, K., & Arthun, C. (2013, October). Predictors of home-based child care providers' participation in professional development workshops and coaching. In *Child & Youth Care Forum* (Vol. 42, pp. 439–455). Springer. https://doi.org/10.1007/s10566-013-9209-y

Sims, S., Fletcher-Wood, H., O'Mara, A., Cottingham, S., Stansfield, C., Van Herwegen, J., & Anders, J. (2021). *What are the characteristics of effective teacher professional development? A systematic review and meta-analysis*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacher-professional-development

Stanton, B. (2015). Teachers' patterns of implementation of an evidence-based intervention and their impact on student outcomes: Results from a nationwide dissemination over 24-months follow-up. *AIDS & Behavior, 19*(10), 1828–1840. https://doi.org/10.1007/s10461-015-1110-2

Strunk, K. O., Marsh, J. A., Bush-Mecenas, S. C., & Duque, M. R. (2016). The best laid plans: An examination of school plan quality and implementation in a school improvement initiative. *Educational Administration Quarterly, 52*(2), 259–309. https://doi.org/10.1177/0013161X15616864

Thoonen, E. E., Sleegers, P. J., Oort, F. J., Peetsma, T. T., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly, 47*(3), 496–536. https://doi.org/10.1177/0013161X11400185

Watson, G. (2006). Technology professional development: Long-term effects on teacher self-efficacy. *Journal of Technology and Teacher Education, 14*(1), 151–165.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). Professional development in the *United States: Trends and challenges: Phase II of a three-phase study: Technical report*. National Staff Development Council.

Weiss, I. R., & Pasley, J. D. (2006). Scaling Up Instructional Improvement through Teacher Professional Development: Insights from the Local Systemic Change Initiative. *CPRE Policy Briefs, 32*, 1–16.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. United States Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Handbooks

Zehetmeier, S., & Krainer, K. (2011). Ways of promoting the sustainability of mathematics teachers' professional development. *ZDM-Mathematics Education, 43*, 875–887.

For more information visit
edresearch.edu.au

171025.SS.AA.V1