

Exploring the shift to an improvement-oriented approach to external evaluation in Aotearoa New Zealand: The case of the Education Review Office

Ro Parsons and Joanna Higgins

Victoria University of Wellington

External evaluation in Aotearoa New Zealand is an important accountability mechanism in education. In 2019 the Tomorrow's Schools Independent Taskforce recommended that the Education Review Office (ERO) develop and implement an improvement-oriented approach to external evaluation in schools. This approach requires fundamental shifts in evaluation practice. In implementing an improvement-oriented approach, while maintaining accountability functions in a public sector context, evaluators need to balance key tensions: relational, epistemological, pedagogical, contextual, political, methodological, and organisational. The role of the evaluator in implementing ERO's new approach, and managing the shifts required, is key to the approach's success. Building evaluation capability and capacity and strengthening the evaluation evidence base are critical areas for further development.

Background

The Education Review Office (ERO) was established in 1989 under the structural reforms of the New Zealand education system. Under those reforms, individual schools became the unit of education administration – “autonomous and self-managing organisations run by boards of trustees under the terms of a charter between the school and the government” (French, 2000, p. 1). The education system became one of the most devolved in the world.

Tomorrow's Schools (Department of Education, 1988) stated that the Review and Audit Agency would be an independent body, accountable through a chief executive officer, to the Minister. Its role was to ensure that institutions were accountable for government funds spent and for meeting the objectives set out in their charter. The agency would also comment on the performance of other elements in the system such as the supply of general advisory services to institutions and the Ministry of Education's provision of policy advice and overseeing of policy implementation. The agency would review institutions using multi-disciplinary teams, ensuring that regular reviews helped boards to meet their objectives and review their own performance. The cycle of review visits was envisaged as two-yearly. The review team would produce a report that identified strengths and weaknesses and made recommendations for improvement. The institution could then make changes in practice, receive a second visit one term later and a final report would be published (Section 2.3).

Both *Administering for Excellence* (Taskforce to Review Education Administration, 1988) and *Tomorrow's Schools* (Department of Education, 1988) stated that review reports on the performance of institutions would be public documents. Under *Tomorrow's Schools* those conducting reviews would “not take a responsibility for advice and guidance to institutions, apart from any recommendations they wish to make in their reports and through their review process” (1988, p. 22). Thrupp and Smith (1999) point to early changes in ERO's name from the Review and Audit Agency (Department of Education, 1988) to the Education Review and Audit Agency in May 1989 and then the Education Review Office in October 1989, as evidence of the organisation's contested beginnings.

The role and function of the Education Review Office

The role and function of ERO has not changed significantly since the Review and Audit Agency was established. Most changes have related to the focus of, and approach to, aspects of its evaluation role (for example, the increased emphasis on equity and excellence) and how functions are operationalised (for example, the composition of review teams, the timing of the review cycle and the nature of reporting).

Under Part 5 of the Education and Training Act 2020, the Chief Review Officer has the power to:

administer reviews, either general or relating to particular matters, of the performance of applicable organisations in relation to the applicable services¹ they provide when directed by the Minister to do so; or on the Chief Review Officer's own motion; and administer the preparation of reports to the Minister on the undertaking and results of the reviews; and give the Minister any other assistance and advice on the performance of the applicable organisations that the Minister requires. (Subpart 3, section 463)

The Chief Review Officer may designate any suitably qualified person a review officer; and must ensure that every person for the time being so designated has a certificate to that effect, in a form approved by the Chief Review Officer (section 465). Review officers have powers of entry and inspection (Part 6, Subpart 6, section 622) and must provide proof of identity before acting under section 622. ERO has an important accountability function in the education system.

Tensions in ERO's role and function

Tensions in ERO's role as an external evaluation agency with a range of evaluation functions have been evident since its establishment. The purposes of evaluation can be described in terms of three general perspectives: evaluation for accountability (the measurement of results or efficiency); evaluation for development (the provision of evaluative help to strengthen institutions); and evaluation for knowledge (the acquisition of a more profound understanding in some specific area or field) (Chelimsky, 1997). These perspectives parallel the three primary uses of evaluation identified by Patton (2008): judging merit or worth, improving programmes, and generating knowledge. Mark, Henry and Julnes (2000) add a category, oversight and compliance, and emphasise the role of evaluation in social betterment.

The core elements of the review agency proposed in *Administering for Excellence* (Taskforce to Review Education Administration, 1988) and *Tomorrow's Schools*

¹ “applicable organisation” means an organisation that provides an applicable service; “applicable service” means an education service to which any of sections 462(1), 463, 464, 465, 622 and 624 apply.

(Department of Education, 1988) are evident in ERO's current role and functions. Although changes have occurred since the agency was established, ERO's role and function in the Aotearoa New Zealand schooling context encompasses both accountability (including compliance with regulatory requirements) and improvement. The Organisation for Economic Cooperation and Development (OECD) review (Nusche et al., 2012) highlighted the challenges for ERO in balancing the tensions inherent in this dual purpose. At a system level ERO's evaluation role includes a knowledge generation function (Patton, 2008) through the work of *Te Ihuwaka*, the Education Evaluation Centre within ERO.

Changes in ERO's approach to external evaluation

Changes in ERO's evaluation approach over time reflect the tensions in the agency's role and function. Government policy settings, specific reviews (Ministry of Education, 1990; Austin et al., 1997; Rodger et al., 2000; Ministry of Education, 2019) and ERO's strategic leadership have been important influences on changes in ERO's evaluation focus and approach. A significant impetus for change has been the demand to improve overall system performance, the quality of education provision and learner outcomes, and strengthen ERO's role in driving that improvement. The following section illustrates that from the outset, challenges associated with ERO's relative emphasis on compliance, accountability and improvement as mechanisms for lifting education quality and effectiveness, as well as notions of differentiation and a flexible response to the evaluation context, were evident.

Early development

ERO's initial approach focused on evaluating each education institution in terms of their charter objectives. In 1991 a new Labour Government and Minister led to a greater emphasis on educational achievement. ERO increased its focus on education outcomes and the dimensions of practice that contributed to school effectiveness. The "Five Headings" approach included: student achievement; learning and teaching; assessment and evaluation; leadership and management; and community participation (French, 2000). During this time Discretionary Audits (follow up visits by ERO after one year) were also undertaken to encourage schools to take action in responding to recommendations.

Assurance audits and effectiveness reviews

The focus on both compliance (through assurance audits) and education performance was challenging and the relative emphasis on each function by review teams varied. From an evaluation perspective, a compliance-oriented approach is relatively straightforward while a focus on performance requires the evaluator to make a professional judgement about quality and effectiveness. Education institutions were also at different points on the development continuum, with some still having low levels of compliance with legislative requirements. ERO separated the focus on compliance and education outcomes to sharpen the value and usefulness of reviews. The focus on outcomes through effectiveness reviews also highlighted the issues associated with the availability, validity and reliability of student outcome data.

Accountability reviews

As understanding and practice related to meeting legal obligations increased across the education sector, the focus of ERO's evaluation approach shifted from compliance to the quality of education provision. The accountability review, introduced in 1997, was

designed to enable review teams to respond flexibly to the context of the education institution and “span a continuum of accountability [from] input management (for example, governance) through output delivery (for example, curriculum delivery) to educational outcomes (for example, student achievement)” (Education Review Office, 1996, p. 18). The external evaluation process included the internal review of compliance issues by institutions. *Achieving Excellence: A Review of External Evaluation Services*, the report of the review set up by the National Government to “recommend ways of increasing the effectiveness of the Education Review Office’s contribution to improving the quality of education outcomes for students in schools and early childhood centres” (Austin et al., 1997, p. 77) was largely supportive of ERO’s accountability orientation and approach to evaluation.

Education reviews

In 2000, however, the new Labour Government initiated another review of ERO. The review committee focused on the role of external evaluation, the relationship between external and internal evaluation, and the link between external evaluation and follow up action and proposed that reviews adopt an “assess and assist” approach. The recommendations in *A Review of the Roles and Responsibilities of the Education Review Office* represented a more deliberate shift from an accountability-oriented approach to an improvement-oriented approach to external evaluation: “The focus of reviews is on educational improvement although the Office maintains compliance functions” (Rodger et al., 2000, p. 2):

ERO’s approach is intended to maximise the impact [of reviews on educational improvement] by focusing on evaluation for both improvement and accountability purposes ... by implementing review processes that focus on improvement [and] continuing to reach independent judgements about the quality of education provided to young New Zealanders. (Education Review Office, 2003, p. C1:8)

A differentiated approach to education reviews

In 2008 the development of a differentiated approach to external evaluation was designed to focus evaluation resources more deliberately on education institutions that were not performing well, and to ensure that external evaluation was more responsive to context. Political drivers included a requirement that compliance demands on schools that were performing well were reduced, and the Government’s broader economic rationale that all public agencies delivered their services smarter, better and for less.

Central features of the differentiated evaluation approach were the integration of external and internal evaluation and a focus on working in partnership with the education sector to build evaluation capability (Mutch, 2012). Where a school’s internal evaluation was limited, ERO’s external evaluation role was proportionally greater and focused on building school performance and evaluation capability. Where a school’s internal evaluation was well developed, school accountabilities were met and there were positive student outcomes, ERO’s external role was proportionally less.

Increasing the focus on equity and excellence

ERO’s introduction of the Accelerating Student Achievement approach in 2016 was a response to the continuing challenge for the Aotearoa New Zealand education system – achieving equity and excellence of education outcomes. The Accelerating Student

Achievement approach focused more sharply on student outcomes – in particular Māori learners, Pacific learners and those with learning support needs – and the acceleration of progress. It reviewed the conditions in schools supporting ongoing improvement in learner outcomes and the identification of priorities for improvement. ERO also increased its evaluation capacity building activity through the provision of workshops and publication of new evaluation resources.

The challenge: Evaluation influence

The need to strengthen the relationship between external evaluation and education improvement has been articulated through external reviews (Rodger et al., 2000) and by ERO as an agency since 2000. Despite the changes in ERO's approach to external evaluation over time, the influence of external and internal evaluation on school improvement in the Aotearoa New Zealand setting, and evaluation capability and capacity across the system, has remained variable.

The final report by the Tomorrow's Schools Independent Taskforce, *Our Schooling Futures: Stronger Together Whiria Ngā Kura Tuātinitini*, (2019) identified limitations in ERO's capacity to influence improvement. While some schools found ERO reviews useful, their event-based nature and infrequency limited effectiveness. The reviews were viewed as high stakes because of the public nature of reporting. Schools were incentivised to put on their 'best face on' rather than focus on improvement discussions. The time spent by reviewers on site, as well as perceptions of their credibility in facilitating improvement and learning opportunities, were also seen as limitations. Principals and teachers wanted to see a stronger emphasis on an ongoing review and connected support process that was relationship based.

Although the Taskforce recommended changes in ERO's functions to focus on system-wide evaluation, the Government did not progress this recommendation because "individual school reviews provide important information for schools, whānau and communities." (Ministry of Education, 2019, p. 15). This decision recognised the importance of ERO's accountability role in the system. The position document *Supporting all schools to succeed* (Ministry of Education, 2019) stated that ERO needed to further develop its approach to external evaluation and "strengthen the capability of schools to undertake self-evaluation and continuous improvement, including ensuring effective engagement with whānau and communities" (p. 15).

The potential for influence: An improvement-oriented approach to external evaluation

Internationally, the evidence-base related to the effects of external evaluation and inspection approaches and explanatory features across different jurisdictions has significant research gaps and the overall picture lacks coherence (Penninckx, 2017). Little research has been undertaken in Aotearoa New Zealand related to school internal evaluation, and the effect of ERO's external evaluation, and impact on student outcomes. Parsons (2006) identified that the extent to which ERO's approach was perceived to assist, and assisted, each school to improve was a complex interaction between the school's review history, evaluator practice, school conditions and participants' responses during the review process. From the perspective of promoting continuous improvement in schools, the evaluation process was as important as the results generated by the evaluation, and the relational dimension of evaluator practice was paramount:

The influence of the review process is mediated by the reviewer. The influence of the review results is mediated by members of the school community, individually and collectively. In assisting schools to improve, ERO's approach to external evaluation therefore has a differential influence in each review context. (Parsons, 2006, p. 326)

ERO's improvement-oriented evaluation approach, *Te Ara Huarau* (Education Review Office, 2022) is a response to the Tomorrow's Schools Independent Taskforce review. *Principles of Practice: Education Evaluation for Improvement in Schools and Early Childhood Services* (Education Review Office, 2021) articulates the evaluation theory and research related to the role of evaluation in supporting and promoting improvement that underpins the approach. From this perspective, *Te Ara Huarau* has the potential to strengthen the influence of external evaluation on continuous school improvement.

External evaluation is positioned as complementing the evaluation activities of schools (MacBeath, 2012) and supporting them in their improvement journeys. The evaluation process is framed as collaborative, dialogic and action oriented (Preskill & Torres, 2000). At the heart of the external evaluation approach is the use of a common set of evaluation indicators, *School Evaluation Indicators: Effective Practice for Improvement and Learner Success* (Education Review Office, 2016), that articulate the evidence base related to what matters most in improving learner outcomes. The process involves participants in the co-construction of evaluation knowledge and understandings so that the use of evaluation findings in decision-making is increased (Adams et al., 2014). The collaborative process (Schulha et al., 2016) supports responsiveness to culture and the context (La France et al., 2012; Cram et al., 2015). Evaluation that focuses on what will make the most difference, is technically rigorous and supports evaluative thinking, reasoning and decision-making strengthens organizational capacity for improvement (Labin et al., 2012; Cousins et al., 2014). The evaluation process is designed to promote both professional accountability and external accountability (Fullan et al., 2015).

The whakatauaiki *Ko te tamaiti te pūtake o te kaupapa: The child – the heart of the matter* articulates the focus of education evaluation in Aotearoa New Zealand. The primary focus of evaluation is the learner, their learning and education outcomes.

Shifting the balance: A significant challenge

Te Ara Huarau (Education Review Office, 2022) is designed to strengthen the relationship between evaluation and continuous improvement. Implementation of the approach involves fundamental changes to how external evaluation is undertaken in the schooling context. Instead of the external evaluation cycle being event-based and time-bound, with the evaluation process carried out by evaluation teams, each external evaluator has responsibility for a portfolio of schools, and the evaluation and administration activities associated with that portfolio. The external evaluator must develop an ongoing evaluation relationship with each school, differentiate the extent of their involvement depending on the school's capacity for ongoing improvement, and manage the tensions between an accountability-oriented approach to external evaluation and an improvement-oriented one.

Evaluation can both legitimate and respond to the system of performance measurement and accountability evident in the public sector, and/or act as a mechanism to broaden democracy by engaging those directly involved and including the voices of the less powerful (Chouinard, 2013). In discussing the case for participatory evaluation in the public sector, Chouinard reminds us of the challenges involved in implementing

participatory approaches to evaluation in contexts where the culture of evaluation is embedded in a culture of accountability. While acknowledging that there can be a mixing of evaluation functions and purposes in the context of practice, as is the case with ERO, Chouinard identifies seven areas of tension that can influence the implementation of participatory approaches to evaluation in public sector contexts: relational, epistemological, pedagogical, contextual, political, methodological, and organisational. It is the evaluator who must navigate and balance these tensions in the context of practice. In any school the evaluator will be operating on an accountability | improvement continuum in relation to the areas of tension identified depending on what the evaluation evidence indicates about the school’s improvement trajectory.

These areas of tension provide a useful framework for examining the potential challenges for ERO in the shift from an accountability-oriented external evaluation approach to an improvement-oriented one. Table 1 uses Chouinard’s (2013) framework to illustrate the tensions between the two approaches.

Table 1

Tensions between accountability and improvement

Area of tension	Accountability orientation	Improvement orientation
Political	Limited participation of stakeholders “One size fits all” approach Mechanism for identifying and responding to poor performance	Wider engagement and participation of parents, family and community Differentiated response Mechanism for improvement and evaluation capacity building
Epistemological	Instrumental/technical knowledge orientation Definition of legitimate knowledge external to context External expert teams make objective evaluation judgements, and identify actions required and recommendations	Social constructivist knowledge orientation Recognition of social, historical and political contexts Giving effect to Treaty of Waitangi partnership and valuing of indigenous knowledges Individual evaluators work with schools to construct evaluation understandings and improvement actions
Organisational	Top-down approach Teams of evaluators and centralised scheduling Evaluation relationships constrained by time allocation, on-site process and resource Hierarchical, bureaucratic relationships	Bottom-up approach Individual evaluators responsible for portfolio of schools Differentiated allocation of evaluation time Collaborative, learning-focused relationships
Contextual	Narrow definition of context Stakeholder engagement minimised Focus on outcomes and centrally determined standards of quality and effectiveness	Responsive and specific to context Sensitive to dynamics of local community Recognises diversity in student and community needs in evaluation process

Methodological	Methodological decisions made by evaluation team within scheduling and resourcing constraints	Methodological decisions differentiated and responsive to contextual conditions
Pedagogical	Evaluation is a technical process Focus on value: merit, worth and significance Evaluation product (the report) and reporting to stakeholders is important Relationships of accountability and compliance Evaluation findings provide evidence of quality and effectiveness, and assist programme decision making and resource allocation	Evaluation is an educational process Focus on facilitating learning and capacity building The process of evaluation and use of findings is important Participatory, collaborative relationships Evaluation process contributes to understanding and use of findings for improvement
Relational	Evaluator role is independent and objective Evaluator relationships are detached, formal and time bound Evaluation relationships focused on meeting accountability requirements, “pleasing ERO and getting a good report”	Evaluator has ongoing relationship with school Relational trust creates the conditions that will enable stakeholder participation and dialogue Evaluation relationships focused on joint construction of evaluation knowledge for improvement

Balancing the tensions: The role of the evaluator

The analysis of the tensions associated with ERO’s implementation of an improvement-oriented approach to external evaluation (see Table 1) highlights some significant shifts in terms of evaluation practice. An improvement-oriented approach requires contextual and cultural responsiveness and sensitivity to the dynamics of the school community. The evaluation process is critical, providing an opportunity to develop collaborative, learning-focused relationships over time and to build evaluation capacity. Methodological decisions need to be differentiated and fit for purpose. The relational dimension is central in facilitating evaluation for improvement over time.

Evaluation of the early implementation of *Te Ara Huarau* (Education Review Office, 2022) emphasises the key role of the evaluator in the success of ERO’s improvement-oriented approach (Goodrick, 2022). The evaluator works alone in the field. From an external evaluation perspective, the approach expands, and increases the complexity and demands of, the evaluator’s role in a range of dimensions of expertise.

Evaluation knowledge and expertise

The knowledge and understanding associated with ERO’s accountability role in the education system, and the legislative and regulatory context in which it operates, remain important. Ultimately ERO evaluators must be able to identify poor performance and recommend external intervention.

In enacting a more improvement-oriented approach to external evaluation, the evaluator also requires deep disciplinary knowledge and understanding of education and evaluation as well as expertise in evaluation practice and evaluation capacity building. An improvement-oriented evaluation approach positions the evaluator as a professional whose practice is characterised by adaptive expertise (Bransford et al., 2000; Mylopoulos & Regehr, 2009). In an education context, adaptive expertise is characterised by self-

efficacy and agency and enhanced by the opportunity to work collaboratively (Timperley, 2013).

Relational expertise

Evaluators need to establish relational trust through their engagement with key participants in each school in a portfolio. The evaluator needs to be able to develop and maintain professional relationships that are learner-focused and challenge the current state. This relational expertise enables robust discussions about evaluation data and information and facilitates openness to change and the development of effective strategies to improve outcomes. A successful evaluation process must facilitate participants' engagement and contribution and promote evaluative thinking and learning as the evaluation progresses. The evaluation process is an important mechanism for building participants' evaluation capability and confidence.

Cultural and contextual expertise

ERO expectations related to giving effect to Te Tiriti o Waitangi partnership (Education Review Office, 2023) also increase the complexity of the role. Community engagement and participation require a strengthened focus on the wider school context and consideration of different cultural values, perspectives and voices. Evaluators need to develop their cultural knowledge and understandings to be able to participate and respond appropriately and confidently in different contexts. An understanding of, and flexible response to, the school context supports the identification of the evaluation focus and the selection of an evaluation approach and methods that are fit for purpose. The evaluator's credibility, interpersonal skills, and ability to effectively engage in dialogue influence stakeholders' receptiveness to evaluation findings and use for improvement (Hofstetter & Alkin, 2003).

Organisational expertise

The professional and organisational responsibilities and demands associated with a portfolio of thirty to forty schools are significant. From an operational perspective, the evaluator must differentiate the allocation of time, depending on where each school is at on the improvement continuum, and maintain ERO's internal quality assurance and reporting procedures for each school. From a professional perspective the evaluator needs to remain engaged in each school's evaluation for improvement process, maintain effective working relationships, and facilitate focused and useful evaluation on-site processes that build evaluation capability over time.

Increasing the influence of evaluation on improvement: What will it take?

External evaluation in Aotearoa New Zealand is an important mechanism for accountability and improvement in education. Since ERO was established in 1989, a series of organisational development initiatives have been designed to shift the focus and implementation of external evaluation to an improvement orientation. ERO's response to the recommendations of the Tomorrow's Schools Independent Taskforce (2019) represents another opportunity to increase the influence of evaluation on school and system improvement. The organisational, operational and resourcing changes being implemented are intended to better support a continuous improvement approach. Increasing the influence of evaluation on improvement requires a sharpened focus on

evaluator knowledge and expertise, and a systematic approach to building evaluation capability and capacity and strengthening the evaluation evidence base.

A focus on evaluator knowledge and expertise

The success of an improvement-oriented evaluation approach is dependent on the external evaluators who work with schools, and the opportunities available to schools to build capacity to do and use evaluation for improvement (Cousins et al., 2014). ERO evaluators must not only have the knowledge and expertise to be able to enact an improvement-oriented approach to evaluation in a range of contexts and facilitate the building of evaluation capability and capacity. They must also have knowledge and expertise related to school effectiveness and improvement and the evidence base about what makes the most difference in improving outcomes in the context of education practice.

Bryk (2015) has highlighted the complexity inherent in schools, especially the increasing heterogeneity of classrooms, the density of organisational activity and its lack of coherence, and the increasing demands in terms of professional knowledge: “Task and organisational complexity beget wide variation in performance” (p. 471). Addressing the wide variability in performance, especially within schools, is the challenge for the Aotearoa New Zealand education system (Schmidt et al., 2015). External professionals who work in school evaluation and improvement, reflect this variability in the knowledge and expertise they bring to their roles and often lack access to high quality professional development opportunities.

Building evaluation capability and capacity

Building evaluation capability and capacity is essential to ensure the future of evaluation “as a professional practice committed to promoting the public good” (Schwandt, 2015, p. 148). There is currently a lack of coherent provision (and associated resourcing) related to education pathways in evaluation, including for example, high quality preparation through university programmes, access to specialist post-graduate qualifications and ongoing professional learning and development opportunities.

It is also important that such provision ensures a balance between evaluation as a technical activity and evaluation as professional practice. Schwandt (2015) articulates a concern that professional learning programmes increasingly focus on developing highly skilled technicians through, for example, training in models, techniques and methods of evaluation. While technical expertise is important it is insufficient in an improvement-oriented evaluation approach where judgements about quality and effectiveness are important in decision-making about action that will make the most difference for equity of education opportunity and outcomes.

Increasing the influence of external evaluation on education improvement, as well as building expertise in internal evaluation (Earl, 2014) and accelerating improvement (Bryk, 2015) in schools will require a coherent, systematic approach to the provision of evidence-based learning and development opportunities for members of the education workforce engaged in capability and capacity building roles.

Strengthening the evaluation evidence base

There is a critical need to strengthen the evaluation evidence base in Aotearoa New Zealand. The research base related to evaluation practice is limited. Internationally, situational awareness and strong interpersonal skills have been identified as features of

the practice of effective and experienced evaluators (Garcia & Stevahn, 2020). There is little research on evaluator competency and training, particularly evaluators' perceptions of their professional learning and development needs. Galport and Azzam (2017) found that practising evaluators identified professional practice and systematic inquiry as the most important competencies for conducting successful evaluations. The relationship between the competencies required in ERO's improvement-oriented evaluation approach and evaluators' training needs requires further investigation.

There have also been calls for more research focused on evaluation capacity building and measuring evaluation capacity within organisations and the intended outcomes. Current findings highlight the potential to under- or over-estimate evaluation capacity depending on who provides the organisational assessment (Preskill, 2014; Fierro & Christie, 2017). Given the emphasis on evaluation capacity building envisaged by the Tomorrow's Schools Independent Taskforce (2019), the relationship between on-site evaluation practice and education improvement in the context of ERO's new approach is an important research focus.

There is no research in Aotearoa New Zealand that compares the effectiveness of individual and team approaches to external evaluation, and the consequent impact on organisational improvement and outcomes. In discussing the training needs for evaluation site visitors, Haynes and Johnson (2017) point out that "one is the loneliest number" (p. 79). Working in pairs or teams increases the quality of evidence gathering and follow-up action, provides powerful learning opportunities, and increases the likelihood of site visit success. The effectiveness and impact of a portfolio-based approach to external evaluation needs investigation.

Finally, Goodrick's (2022) evaluation case studies of schools in the early phase of Te Ara Huarau process identified some key questions and recommendations related to generating a fuller understanding of the approach during implementation: managing school expectation within resourcing constraints; extending evaluative capability across schools and within classrooms; aligning ERO's work with other partners; building and maintaining internal capability in evaluation; and supporting schools in internal evaluation and external accountability. As well as generating an understanding of what is working (and what is not) and opportunities for further improvement, future external evaluation(s) should focus on the effectiveness of the approach and its impact on outcomes at school and system level.

The challenge of system change

ERO's implementation of an improvement-oriented approach to external evaluation is only one component of a suite of education policy initiatives and change programmes aimed at addressing the decline in education performance and improving the equity and excellence of student outcomes. Since the introduction of *Tomorrow's Schools* in 1989 the roles and functions of education agencies have shifted so that there is currently overlap in their evaluation, research and improvement functions. There is also fragmentation and variability in the approach to system change and the design and implementation of improvement activities.

The synthesis and articulation of evidence linked to outcomes has been undertaken through initiatives such as the Ministry of Education's *Iterative Best Evidence Synthesis Hei Kete Raukura* programme and ERO's development of *School Evaluation Indicators: Effective Practice for Improvement and Learner Success* (Education Review Office, 2016). However, there is no coherent, systematic approach to using and further developing what

we know about what works and evaluating impact. We have examples of successful interventions such as the *Learning Schools Model* (Lai & McNaughton, 2016) that draw on design-based research (Anderson & Shattuck, 2012) and improvement science, but we have yet to use this evidence in a systematic way to design and resource system implementation approaches.

Our success will depend on our ability to work together and accelerate how we learn to improve. If the system is to achieve its ambition of improving the equity and excellence in outcomes for all learners, agencies need to work collaboratively to increase the consistency of evidence informed practice and systematically build capability at every level of the system.

References

- Adams, A., Nnawulezi, N., & Vandenberg, L. (2014). "Expectations to Change" (E2C): A participatory method for facilitating stakeholder engagement with evaluation findings. *American Journal of Evaluation*, 36(2), 243-255.
<https://doi.org/10.1177/1098214014553787>
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16-25.
<https://doi.org/10.3102/0013189X11428813>
- Austin, M., Parata-Blane, A., & Edwards, W. (1997). *Achieving excellence: A review of the education external evaluation services*. Ministry of Education.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.
- Bryk, A. S. (2015). Accelerating how we learn to improve. *Educational Researcher*, 44(9), 467-477. <https://doi.org/10.3102/0013189X15621543>
- Chelimsky, E. (1997). The coming transformations in evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 1-26). Sage.
- Chouinard, J. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation* 34(2), 237-253.
<https://doi.org/10.1177/1098214013478142>
- Cousins, B., Goh, S., Elliott, C., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In J. B. Cousins & I. Bourgeois (Eds.), *Special issue: Organizational capacity to do and use evaluation. New Directions for Evaluation*, 141, 7-23.
<https://doi.org/10.1002/ev.20076>
- Cram, F., Kennedy, V., Paipa, K., Pipi, K., & Wehipeihana, N. (2015). Being culturally responsive through kaupapa Māori evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*, (pp. 289-311). University of Auckland.

Department of Education. (1988). *Tomorrow's Schools: The Reform of Education Administration in New Zealand*. Government Printer.

Earl, L. (2014). *Effective school review: Considerations in the framing, definition, identification and selection of indicators of education quality and their potential use in evaluation in the school setting*. Background paper prepared for the review of the Education Review Office's Evaluation Indicators for School Reviews.

Education Review Office. (1996). *Annual Report: 1 July 1996 to 30 June 1997*. Education Review Office.

Education Review Office. (2003). *Manual of standard procedures for education reviews*. Education Review Office.

Education Review Office. (2016). *School evaluation indicators: Effective practice for improvement and learner success*. Education Review Office.
<https://ero.govt.nz/how-ero-reviews/schoolskura-english-medium/school-evaluation-indicators>

Education Review Office. (2021). *Principles of practice: Education evaluation for improvement in schools and early childhood centres*. Education Review Office.
<https://ero.govt.nz/sites/default/files/2021-11/Principles%20of%20Practicev8.pdf>

Education Review Office. (2022). *Te ara huarau: Your go-to-guide*. Education Review Office. <https://ero.govt.nz/sites/default/files/2022-04/Te%20Ara%20Huarau%20Go%20To%20Guide%20March%202022.pdf>

Education Review Office. (2023). *Strategic intentions 2023-2026*. Education Review Office. <https://ero.govt.nz/our-strategies/strategic-intentions-2023-2026>

Fierro, L., & Christie, C. (2017). Evaluator and programme manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation*, 38(3), 376-392. <https://doi.org/10.1177/1098214016667581>

French, A. (2000). *The heart of the matter: How the Education Review Office evaluates pre-tertiary education*. Cases in public sector innovation; No. 9. Victoria Link Ltd. Victoria University of Wellington.

Fullan, M., Rincon-Gallardo, S., & Hargreaves, A. (2015). Professional capital as accountability. *Education Policy Analysis Archives*, 23(15), 1-22.
<https://doi.org/10.14507/epaa.v23.1998>

Galport, N., & Azzam, T. (2017). Evaluator training needs and competencies: A gap analysis. *American Journal of Evaluation*, 38(1), 80-100.
<https://doi.org/10.1177/1098214016643183>

Garcia, G., & Stevahn, L. (2020). Situational awareness and interpersonal competence as evaluator competencies. *American Journal of Evaluation*, 41(1), 107-124.
<https://doi.org/10.1177/1098214018814941>

Goodrick, D. (2022). *Schools evaluation for improvement approach: Implementation case studies*.
<https://ero.govt.nz/how-ero-reviews/how-ero-reviews-schoolskura-english-medium/te-ara-huarau-the-new-approach-to-evaluation/feedback-on-the-new-schools-operating-model>

Haynes, M., & Johnson, A. (2017). Training needs of site visitors. In R. K. Nelson, & D. L. Roseland (Eds.), *Conducting and using evaluative site visits*, 156 (pp. 75-82). Jossey-Bass and American Evaluation Association. <https://doi.org/10.1002/ev.20268>

Hofstetter, C., & Alkin, M. (2003). Evaluation use revisited. In T. Kellaghan, & D. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 197-222). Springer.

Labin, S., Duffy, J., Meyers, D., Wandersman, A., & Lesesne, C. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307-338. <https://doi.org/10.1177/1098214011434608>

LaFrance, J., Nichols, R., & Kirkhart, K. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conner (Eds.), *Context: A framework for its influence on evaluation practice*, (pp. 59-74).
<https://doi.org/10.1002/ev.20027>

Lai, M. K., & McNaughton, S. (2016). The impact of data use professional development on student achievement. *Teaching and Teacher Education*, 60, 434-443.
<https://doi.org/10.1016/j.tate.2016.07.005>

MacBeath, J. (2012). *Future of teaching profession*. University of Cambridge.

Mark, M., Henry, G., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programmes*. Jossey-Bass.

Ministry of Education. (1990). *Today's Schools: A review of the education reform implementation process*. Report prepared for the Minister of Education, April, 1990. Ministry of Education.

Ministry of Education. (2019). *Supporting all schools to succeed. Reform of the Tomorrow's Schools system*. Ministry of Education.
<https://conversation.education.govt.nz/assets/TSR/November-2019/TSR-Government-Response-WEB.pdf>

Mutch, C. (2012). Complementary Evaluation: The development of a conceptual framework to integrate external and internal evaluation in the New Zealand school context. *Policy Futures in Education*, 10(5), 569-586.
<https://doi.org/10.2304/pfie.2012.10.5.569>

Mylopoulos, M., & Regehr, G. (2009). How student models of expertise and innovation impact the development of adaptive expertise in medicine. *Medical Education*, 43(2), 12-132. <https://doi.org/10.1111/j.1365-2923.2008.03254.x>

Nusche, D., Laveault, D., MacBeath, J., & Santiago, P. (2012). *OECD reviews of evaluation and assessment in education: New Zealand 2011*. OECD Publishing.

Parsons, R. (2006). External evaluation in New Zealand schooling (Unpublished doctoral thesis). Massey University.

Patton, M. Q. (2008). *Utilization-focused evaluation (4th ed.)*. Sage.

Penninckx, M. (2017). Effects and side effects of school inspections: A general framework. *Studies in Educational Evaluation*, 52, 1-11.
<https://doi.org/10.1016/j.stueduc.2016.06.006>

Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, 35(1), 116-119.
<https://doi.org/10.1177/1098214013499439>

Preskill, H., & Torres, R. (2000). The learning dimension of evaluation use. In V. Caracelli, & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for Evaluation*, 88, (pp. 25-37). <https://doi.org/10.1002/ev.1189>

Rodger, S., Holden, J., Meade, A., Millar, A., & Smith, B. (2000). *Report to the Minister of Education: A review of the roles and responsibilities of the Education Review Office*. Ministry of Education.

Schmidt, W., Burroughs, N., Ziodo, P., & Houang, R. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher*, 44(7), 371-386. <https://doi.org/10.3102/0013189X15603982>

Schulha, L., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 93-215.
<https://doi.org/10.1177/1098214015615230>

Schwandt, T. (2015) *Evaluation Foundations Revisited : Cultivating a Life of the Mind for Practice*, Stanford University Press, 2015. ProQuest Ebook Central,
<https://ebookcentral.proquest.com/lib/vuw/detail.action?docID=2038803>.

Taskforce to Review Education Administration. (1988). *Administering for Excellence: Effective Administration in Education: Report of the Taskforce to Review Education Administration*. The Taskforce.

Thrupp, M., & Smith, R. (1999). A decade of ERO. In M. Thrupp (Ed.), *A Decade of Reform in New Zealand Education: Where to Now? Special Issue of The New Zealand Journal of Educational Studies*, 186-198.

Timperley, H. (2013). *Learning to Practise: A paper for discussion*. Ministry of Education. <http://hdl.voced.edu.au/10707/257067>

Tomorrow's Schools Independent Taskforce. (2019). *Our schooling futures: Stronger together: Whiria ngā kura tuātinitini*. Ministry of Education. <https://conversation.education.govt.nz/assets/TSR/Tomorrows-Schools-Review-Report-13Dec2018.PDF>

Ro Parsons has had a wide-ranging career in education and has held senior leadership roles in the Ministry of Education and Education Review Office. Ro's academic and research interests are focused on the field of evaluation. Ro's publications relate to mathematics education, teacher professional learning and development, system change and improvement, and evaluation.

<https://orcid.org/0000-0002-3164-615X>

Joanna Higgins is an associate professor at the Faculty of Education, Victoria University of Wellington, New Zealand where she supervises masters and doctoral projects that focus on policy implementation, professional learning and development, leadership, and wellbeing initiatives in schools and universities. Her research from sociocultural perspectives includes instructional reform in mathematics education, professional inquiry of teachers and teacher educators, and analyses of wellbeing policy and practice.

<https://orcid.org/0000-0001-6225-2439>