



TEACHING & LEARNING
RESEARCH INITIATIVE

NĀU I WHATU TE KĀKAHU, HE TĀNIKO TAKU

Observing assessment for learning (AfL) in action: Piloting an observation schedule to inform teacher assessment learning and research

Mary Hill, Helen Dixon, and Eleanor Hawe

May 2020



1. Introduction

In New Zealand we lack valid and reliable evidence-based schedules with which to observe teachers' assessment for learning (AfL) practices. To date, classroom assessment observation schedules have mainly been derived theoretically, used in developing countries with conditions different from New Zealand classrooms (Kanjee & Hopfenbeck, personal communications), or restricted to particular curriculum areas such as science and mathematics (Ruiz-Primo, 2017), literacy (Parr & Hawe, 2009), and writing (Parr & Gadd, 2016). Some approaches follow particular interpretations of formative assessment and AfL (used synonymously in this report), such as those developed for professional development initiatives (e.g., Wiliam & Leahy, 2015). In this project we piloted and modified an observation schedule, and evaluated its usefulness and suitability in the New Zealand primary school teaching context. This schedule, which is called the Developing and Evaluation Measures of Formative Assessment Practice (DEMFAP) observation protocol, was developed in the US from comprehensive video evidence of classroom assessment.

Research evidence strongly supports embedding AfL within classroom teaching as critical to promoting “the powerful and agentic learning that is meant to be at the heart of our national curriculum” (Darr, 2018, p. 1). As summarised in the overview of the OECD/CERI international conference *Learning in the 21st century: Research, innovation and policy* (2008), we now have the evidence that, implemented effectively, AfL promotes the goal of life-long learning and raises achievement, particularly for those most at risk of underachievement. In this way, AfL “has a part to play in bringing about greater equity in student outcomes and supports improved learning to learn skills” (OECD/CERI, 2008, p. 1). A range of projects in New Zealand have investigated ways to increase the formative and AfL practice of teachers (e.g., Poskitt & Taylor, 2009; Hill et al., 2013) in order that we build assessment capable teachers and students (Booth et al., 2016). From the second half of 2020 schools will be able to source funds for professional development in AfL from the Ministry of Education—see priority 1.¹ However, to date, we do not have valid and useful observation schedules with which to capture and inform the improvement of AfL practice. The aim of this project was to bring together teacher and university partners to implement, adapt, and evaluate the DEMFAP or a modified version for possible use in New Zealand primary classrooms.

In a time where our students are increasingly expected to manage themselves as learners (Education Review Office, 2015; Ministry of Education, 2011) in innovative learning environments and work with others collaboratively and online to give peer feedback, teachers use of AfL to bring about such assessment capability in their students is critically important. Such teachers are referred to here as *assessment capable*. The features of AfL have been described and affirmed by several authors over the past two decades (e.g., James & Pedder 2006; Klenowski, 2009; Swaffield, 2011; Wiliam & Leahy, 2015). In short, AfL is a learning and teaching process incorporating five strategies, where teachers and students:

- share goals and criteria to promote student understanding about the goal(s) of learning and what quality performance looks like
- engineer effective discussion and activities that elicit evidence of learning; generate feedback (both external and internal) that moves learning forward
- activate students as learning resources for one another including peer review and feedback; and
- strive for student ownership over, and responsibility for, their learning.

Through these strategies, assessment capable teachers are able to use their curricula, pedagogical, and subject-matter knowledge to notice, recognise, and respond to students' learning needs as they arise, and perhaps more importantly, to “encourage students to be deeply accountable for their own progress and support them to become motivated, effective, self-regulating learners” (Absolum et al., 2009, p. 24). But in order to assist teachers to be assessment capable and to support them to enhance such capability, it is necessary to both understand what assessment capable teachers know and do, and capture evidence of their formative

¹ <https://conversation.education.govt.nz/conversations/curriculum-progress-and-achievement/national-priorities-for-professional-learning-and-development/>

assessment / AfL practices. With this information it is possible for teachers to improve their assessment practice through targeted self and peer feedback, and discussion about “where to next” (Ruiz-Primo, 2017). However, as yet it is unclear to what extent and how observation evidence assists teachers to motivate and activate students to develop evaluative capability. Drawing on existing literature, Booth et al. (2016) provided a conceptual framework that describes ways to identify assessment capable teachers (see Figure 1).

FIGURE 1: Teacher Assessment Capability Framework (reproduced from Booth et al., 2016)

Conditions for effective assessment (Sadler, 1989)	Assessment capable teachers ...
The assessment capable teacher helps students to understand what constitutes quality	<ul style="list-style-type: none"> • authentically share their understanding of quality with students and provide focused feedback about students’ work • adapt teacher resources for a student audience by deconstructing criteria and descriptors, interpreting what they mean, and applying them to real examples of work • explicitly teach students how to access and utilise materials that detail criteria and exemplify quality • model how to judge performance against success criteria or assessment criteria.
The assessment capable teacher helps students develop the metacognitive skills to evaluate their work	<ul style="list-style-type: none"> • create a safe pedagogical, learning-focused environment, where mistakes are seen as opportunities for growth and students are enabled to take responsibility for themselves, as learners• Share with their students their teacher-knowledge about the skills, strategies, and resources needed to carry out a task effectively • devote time, support, and opportunities, in the context of learning, to help students plan, problem solve, and evaluate • explicitly teach students to review and evaluate their abilities, knowledge states and cognitive strategies • explicitly teach self-management skills • explicitly teach students how to self- and peer-assess and how to give and act on feedback • provide sustained and supported experiences in discussing/questioning and improving their work • give students the specific language they might need in order to describe, discuss, and evaluate their learning • model effective problem-solving approaches and are willing and able to be learners themselves
The assessment capable teacher helps students learn strategies to modify their own work	<ul style="list-style-type: none"> • help students to learn how to monitor and improve the quality of their work both during and after its production • provide a variety of exemplars which illustrate what is expected of the students • give explicit teaching of fix-up/improvement strategies • provide time, opportunities, and encouragement within the school day to improve work during its construction • help students to identify where and when to make improvements • provide opportunities for evaluative conversations.

Rather than seek to develop another observation schedule, we piloted, adapted, and tailored the DEMFAP to gather information about a small sample of New Zealand primary school teachers’ AfL practices. As noted above, the DEMFAP was developed for use in the US from the results of a major research project funded by the Institute of Education Sciences (Ruiz-Primo et al., 2016). Following an in-depth analysis of the existing literature, Ruiz-Primo and colleagues defined the construct of formative assessment. Consistent with the New Zealand context, and definitions (Bell & Cowie, 2001; Booth et al., 2016), the formative assessment process underpinning Ruiz-Primo et al.’s (2016) research prioritises students and teachers: clarifying learning goals; eliciting information about learning; interpreting the information; and using the information to adjust instructional processes and interactions (for the full model, see Ruiz-Primo et al., 2016) and includes observations of student involvement in self-reflection. Based upon more than 400 video hours of lessons in 26 science and mathematics

units, cluster analyses differentiated practices capable of distinguishing formative practices among teachers. These practices underpin the DEMFAP which has been trialled with 101 teachers in classrooms in California, Colorado, and Washington State. Given the potential compatibility of DEMFAP with the New Zealand context of AfL, we collaborated with Associate Professor Ruiz-Primo to use the DEMFAP under licence in order to investigate its relevance and appropriateness for use in the New Zealand context. Associate Professor Ruiz-Primo also supported us as an external adviser during this project.

Thus, in this project we used DEMFAP to investigate whether and how this observation schedule could inform New Zealand primary teachers' AfL practice, provide a platform for professional learning and development, and potentially be useful as a research schedule in the New Zealand context. The following research questions guided the project:

RQ1: How well does the existing DEMFAP capture New Zealand primary teachers' AfL practices?

RQ2: What modifications and revisions are necessary to make the observation schedule useful and relevant in the New Zealand primary school context?

RQ3: How effective is the modified observation schedule at capturing critical aspects of the teachers' practice for research purposes?

RQ4: What are teachers' perceptions about the efficacy and usefulness of the modified observation schedule for improving their assessment capability?

2. Research design, data collection, and analysis

The overall aim of this project was to explore the relevance, fit, and usefulness of an existing formative assessment observation schedule for use in the New Zealand primary school context.

Selection of the observation schedule

Three criteria informed our selection of the observation schedule. Firstly, we wanted a schedule that encapsulated the ideas and behaviours that comprised quality AfL practice. Secondly, given the complexity inherent in AfL practice, we wanted to avoid checklists or rating scales that were either overly technical and prescriptive in nature, or focused on a closed set of actions and behaviours. Thirdly, we did not want a schedule that was so descriptive and open ended in nature that an observer would not be clear about the nature of phenomena to be observed. Using the three criteria to inform our decision making, the DEMFAP classroom observation schedule was deemed suitable.

The participants

Consistent with the TLRI requirement that projects must be a collaborative venture between practitioners and researchers, two Auckland primary school principals known to have an interest in and commitment to the development of teachers' AfL practice were approached to be partners and named investigators in the project. Once the project gained funding, a call for the teacher participants in each of the schools was made. To avoid any possible perceived coercion of teachers, the university researchers, rather than the principals, made the call for volunteers. At each school, two of the researchers attended a meeting where the aims of the project were outlined and the expectations of teacher-researcher participants were clarified. At these meetings, the researchers also handed out participant information schedules and consent forms. As can be seen in Table 1, four teachers from Totara School volunteered to participate, whereas in Rimu School, six teachers offered. The teacher participants taught across a range of class levels from Year 1 to Year 6, thus representing the levels typically taught at New Zealand contributing primary schools.

TABLE 1: Demographic information—Teacher participants

Pseudonym	Year level	School
Tch 1	1	Totara
Tch 2	3	Totara
Tch 3	6	Totara
Tch 4	5/6	Totara
Tch 1	3/4	Rimu
Tch 2	2/3	Rimu
Tch 3	2/3	Rimu
Tch 4	2/3	Rimu
Tch 5	2/3	Rimu
Tch 6	1	Rimu

Data collection

Three cycles of data collection and analysis took place during the year. During Cycle 1, the university researchers spent time familiarising teachers with the observation schedule and ensuring that they knew how to use it. The DEMFAP was used and adapted following feedback from teachers and researchers. Cycle 2 involved further adaptation of the schedule based on teachers' feedback and use. Finally, in Cycle 3 further modifications were made to the schedule and an overall evaluation made. During each cycle, teachers worked in pairs to undertake three observations of their colleague using the schedule. Teachers were free to decide which subject area they wished their colleague to observe. Each observation took approximately 30 minutes. At Totara School, these pairs were fluid, with the pairs changing during the project enabling all four teachers to observe each other's practice. At Rimu, the pairs remained constant which resulted in teachers observing in one class only.

During these iterative cycles of use and review of the schedule, data were collected in the following ways (see Table 2). During Cycles 1 and 2, video recordings were taken of each lesson to ensure there was a permanent record of practice. Following each observation each pair discussed the data collected at a time convenient to them. Further, after each observation the observer was debriefed by a university researcher, which took approximately 15 minutes. During these debriefings the observer was encouraged to talk through what they had recorded and to explain how they had used the schedule. They reflected on its usefulness, highlighted any problems encountered, and made suggestions for modification. Hard copies of the completed schedules were collected to see how they had been used and what modifications teachers had made in order to capture a peer's practice.

At the completion of Cycles 1 and 2, two whole-day project meetings involving principals and teachers from the two schools and the three university researchers were convened. A key task was an ongoing appraisal of the observation schedule. At the first whole-day project meeting, to provide a valid point of reference against which an appraisal of the schedule could be made, teachers' attention was drawn to the five strategies that comprise AfL and what they may look like in practice. During this discussion they were asked to reflect on how well the schedule captured the various strategies. Following this discussion, and using the completed observation schedules gathered during Cycle 1, teachers made an appraisal of the schedule's usefulness and "marked up" any possible modifications to the schedule. Furthermore, given that teachers had used the schedule across various essential learning areas, they were asked to make comment on the "fit" of the schedule with a specific learning area. A similar exercise occurred during the second meeting when teachers were asked to evaluate the modified schedule and make further recommendations for amendments.

Near the end of Cycle 2 each teacher participated in a 30–40-minute individual semi-structured interview with a university researcher. The purpose of these interviews was to understand more from the teachers’ perspectives about the efficiency, effectiveness, dependability, and practicality of the observation schedule. Debriefing sessions with teachers, the whole-day project meetings and the semi-structured interviews were all audio-taped and transcribed. Finally, at the end of the project the teachers completed a written evaluation. In this evaluation, teachers were asked to reflect on their personal professional learning from the project, provide further suggestions for modification of the schedule, and make recommendations about the nature and scope of the professional development opportunities needed if the schedule was to be used for teacher professional learning purposes.

TABLE 2: Data collection

Cycle	Data collection
Cycle 1 February–June	<ul style="list-style-type: none"> • Round 1 observations (three per teacher) • Collection of completed observation schedules • Pairs peer debriefing (PD) • Observer debriefing with university researcher • Whole-day project meeting (PM 1) and modification of the existing schedule
Cycle 2 July–September	<ul style="list-style-type: none"> • Round 2 observations with modified observation schedule (three per teacher) • Collection of completed observation schedules • Pairs peer debriefing (PD) • Observer debriefing with university researcher • Individual semi-structured interviews (Int) with 10 teacher-researcher participants • Whole-day project meeting (PM 2) and further modifications to the schedule
Cycle 3 October–November	<ul style="list-style-type: none"> • Round 3 observations with modified observation schedule (three per teacher) • Collection of completed observation schedules • Pairs peer debriefing (PD) • Observer debriefing with university researcher • Written evaluation of the project

Data analysis

In order to answer the first two research questions regarding how well the DEMFAP suits the New Zealand context and the modifications necessary to improve its fit, data from all sources referred to in Table 2 were analysed through collating the information provided by the teachers. The debriefing data, interview transcripts, completed observation schedules, and project meeting data were analysed by identifying common themes across these sources of information to identify the usability of each iteration of the observation schedule and how well it captured the five AfL strategies. In collaboration with our international expert, the observation schedules were then modified for Cycle 2 and following a second round of analysis they were modified again for Cycle 3.

In order to answer RQ3 regarding the usefulness of the DEMFAP for classroom research purposes in New Zealand, the data from the completed observation schedules themselves were collated for every aspect of the Cycle 3 observation schedules to understand how teachers and students were involved in AfL strategies during the observed lessons. This process involved a careful counting of the data recorded on every observation record sheet to form a picture of what took place in the lessons observed. Thematic analysis (Braun & Clarke, 2013) of the debriefing, project meeting and interview transcripts, project meeting materials and written evaluations was conducted to answer RQ4 which investigated teachers’ perceptions about the efficacy and usefulness of the modified observation schedule for improving their assessment capability.

3. Key findings

Research question 1: *How well does the existing DEMFAP capture New Zealand primary teachers' AfL practices?*

As explained briefly above, the DEMFAP classroom observation schedule was selected for use in this project because it was deemed to be both practically orientated and research validated. Specifically, it was considered suitable in that the observation schedules incorporated the five strategies currently seen as integral to AfL practice. Despite the fact the schedule had been developed within the US context, we felt it had the potential for use in New Zealand schools. Given that the schedule was comprised of the five strategies that have been promoted to teachers as part of effective AfL practice, we assumed that teachers would have a degree of familiarity with these. While we anticipated that some modifications would need to be made to the schedule, we were surprised at the extent to which the teachers struggled to use the schedule, particularly during Cycles 1 and 2.

During Cycle 1 most of the teachers struggled to use the schedule to capture AfL practice. For a number of reasons, they did not find the schedule user friendly. In the first instance, teachers found the schedule “very wordy” (Tch 3 RS²) which in turn led them to feel confused about what they were looking for when they were observing a colleague. This situation was exacerbated by the fact that some of the examples (and terms) used within the schedule to illustrate aspects of practice were not commonplace practice in New Zealand. For example, the schedule contained the statements “What happened in the review of the assignment?” and “How did the teacher refer to the products or assessments students need to complete by the end of the day?” (Observation schedule, Cycle 1). These were seen as inappropriate to describe what was happening in these classrooms. As a result, many of the teachers reported that they focused more on the schedule, taking time to see “where something they observed fitted onto the schedule” (Tch 3, TS) than on what was happening within a lesson. In addition to the schedule being too wordy the majority of teachers felt the language used didn’t “quite fit with what we are doing” (Tch 3, TS). Generally, it was felt that it was “very secondary schoolish” (Tch 3, TS) with many teachers mentioning that the way in which practice was described in the schedule “doesn’t always work that way in primary school” (Tch 1, RS). As teacher 3 at Rimu School explained “the language was not really conducive to our school day and our school activities”. In particular, the option of only making comments in relation to whole-class teaching or individual instruction failed to recognise the prevalence of small-group instruction in primary classrooms. Seemingly, the more junior the class, the more apparent it became that teachers found it challenging to “see” a match between the practices described in the schedule and those they were observing in their class. Teacher 6 at Rimu School drew attention to this lack of congruence when she suggested that a modified schedule was needed to describe early years practice:

... it was harder applying some of this to the younger years because some of the examples or some of the ways that things are phrased would never apply in a new entrants classroom ... you need another one for junior kids. (Int.)

The applicability of the schedule’s use across a range of curriculum areas was also mentioned by some teachers. At Rimu School, teacher 5 (Int.) found the schedule too “mathematical and science based” given that a number of the illustrations of practice in the DEMFAP were science or maths related. For example, using phrases such as “explain why you think this is a solution versus a mixture” or “why do you need to first reduce fractions” to illustrate how teachers might ask questions to probe students’ thinking (Observation schedule, Cycle 1) were not seen as helpful as she observed a colleague teaching a writing lesson. Essentially, she could not “find anything related to writing” included in the observation schedule. As a consequence, she felt the schedule needed “to be worded in a more general teaching sense” if it were to be used in a range of areas (Int.).

2 Quotations are identified by teacher (Tch) number at Totara School (TS) and Rimu School (RS).

Following the completion of Cycle 1, modifications to the schedule were made. These affected most aspects of the schedule and continued to be made through to the end of Cycle 3. Ongoing attention was paid to making the language used in the schedule more reflective of New Zealand classroom practice. As a result, language perceived by teachers as associated with secondary schooling or a more formal, structured approach to teaching, learning, and assessment was removed. Terms such as grading, judging, assignments, and testing were omitted. Also, in response to teachers' suggestions, illustrations of practice were modified. For example, the phrase "how did the teacher refer to the *products or assessment* students need to complete by the end of the day" was modified to read "how did the teacher refer to *the tasks* students need to complete" (Observation schedule, Cycle 3). In regard to student actions to be observed, phrases such as "*corrected* their own work/ independently in pairs/small groups" was changed to "*revising* their own work using success criteria/rubric/ independently" (Observation schedule, Cycle 3).

To better reflect New Zealand classroom practice, observers also had the opportunity to record information that pertained to small-group instruction, something that was missing from the original DEMFAP. In response to the request for the observation schedule to have "less words and more white space ... and maybe have a tick box and a space for comments" (Tch 1 RS) the final iteration of the schedule allowed teachers to record additional narrative comments related to what they were observing. Recording this commentary was seen as important to serve as an aide-memoire when teachers were peer-debriefing following an observation.

Overall, as a consequence of the modifications made, the Cycle 3 schedule had less text for the teachers to decipher, used more appropriate language for the New Zealand primary context, and moved towards a combination of checklist items and narrative commentary. These changes were seen as positive amendments by the majority of teachers in that they made the schedule more user-friendly and relevant to their context. Teacher 1 at Totara School encapsulated this appreciation as she explained "I love[d] that the language has become more kiwified ... I love [that] there [was] not so much information on the sheet ... so it makes it easier to read at speed [about] what you need to be doing." (Int.)

Research question 2: *What modifications and revisions are necessary to make the observation schedule useful and relevant in the New Zealand primary school context?*

The modifications and revisions that occurred during the three cycles of implementation resulted in an observation schedule that better reflected and illustrated AfL classroom practice in New Zealand. As a result, the observation schedule has the potential to capture New Zealand primary teachers' practice. However, as we worked with teachers it became apparent that the tool alone would be insufficient if it were to be used for professional learning purposes focused on increasing AfL capability. During the first whole-project meeting it was noticeable that a number of the teachers in the project, especially those from one of the schools, had an incomplete understanding of AfL. This lack of understanding was also confirmed in the individual semi-structured interviews. Typical of a number of teachers, teacher 4 at Rimu School (Int.) believed her lack of knowledge had hampered her ability to use the tool. She felt that if she had had "a little bit of AfL knowledge" she would have been able to use the tool more effectively.

To this end, teachers were clear that if the tool was to be useful and relevant for their teaching colleagues to both capture and improve their practice, its use had to be supported by a programme of professional learning. Acknowledging that teachers would fall along a continuum of AfL understanding and practice, teachers in this project felt that prior to any use of the tool it would be important to determine what teachers knew about AfL and to then plan a programme suited to their identified needs. When asked what a professional learning programme should focus on, teachers were forthcoming in regard to the knowledge and understanding they needed. Specifically, those from Rimu School acknowledged they needed help to understand "what AfL is" through the identification of "the strategies of AfL" (PM 1). While Totara teachers were further along the continuum of understanding and some of these teachers were more confident in aspects of their AfL practice, they still realised that there were gaps in their knowledge and understanding of the five strategies. Teacher 2

at Totara School (Int.), for example, reported while “I am a kind of AfL person” she still felt “insecure” about her AfL knowledge given she had not “had any professional development around it”. Notwithstanding their level of knowledge and understanding, teachers from both schools reported they had often struggled to translate what they saw during an observation and match that to a particular AfL strategy as it appeared in the observation schedule. As a result of this insight, teachers felt they needed more assistance to “recognise what AfL looks like in practice” (PM 1).

One of the ways in which teachers thought their knowledge and understanding of AfL practice could be enhanced was by “having it [practice] modelled” so that teachers as learners could “observe it happening” (PM 1). There was, however, the acknowledgement that observing alone was inadequate to enhance practice. Of importance was the need to “break down the components” (PM 1) and have ongoing conversations about what these might look across a range of subjects and contexts. Overwhelmingly, teachers in the current project felt that observing their colleagues had been beneficial to them because it had provided them with insights related to their practice and offered alternative ways of working.

Research question 3: *How effective is the modified observation schedule at capturing critical aspects of the teachers’ practice for research purposes?*

In order to answer this research question, copies of the completed observation schedules were collected from each teacher following each observation. In Cycles 1 and 2, lessons that were observed were also video recorded. This provided the opportunity to validate a sample of the recorded observations against the lesson recordings. In addition, during the lesson debriefings, what was recorded on each observation sheet was discussed and checked with the observed teacher. Retrospectively, the observation schedules from each teacher were collated and analysis across the items on the observation schedules carried out.

For each of the five AfL strategies—clarifying, sharing goals/criteria, and understanding what quality looks like; eliciting evidence about learning through effective dialogue/activities; providing feedback to move learning ahead; activating students as resources for one another; and activating students as owners of their own learning—data from the observation schedules were collated to form a picture of teachers’ practices in the lessons observed. These collations demonstrated that, by using the information on the observation schedules, it was possible to see how teachers used/didn’t use these AfL strategies within the observed lessons and how well the observation schedule represented data about each of the strategies. This analysis was more successful for some of the strategies than others, due to the items included on the observation schedule. In terms of clarifying goals and sharing what quality performance looks like (AfL Strategy 1), the data across the approximately 25 lessons observed and recorded indicated that two teachers were not observed clarifying learning goals at all, three briefly mentioned the tasks and activities they expected students to complete, and five were observed sharing goals, elaborating on criteria, and/or sharing examples of what quality looked like during three of the lessons observed.

Regarding eliciting information about learning (AfL Strategy 2), almost all lessons observed were conducted with either a whole class or in small groups. In approximately 75% of the lessons observed, teachers mainly asked how and why questions, content or fact-based questions or asked how or why the students did something. Students’ explanations were sought in almost every lesson observed. In contrast, very few lessons incorporated going over something students had done previously, suggesting that these classrooms were dialogically interactive rather than focused on autonomous student work. In fact, the teachers advised that the “homework” and “assignment” categories should be removed from the observation schedule (PM 1). That said, teachers elicited information in recorded form from students in about half the lessons observed. In terms of providing feedback (AfL Strategy 3), there was evidence that all teachers provided useful information, did some reteaching, helped students develop strategies, and summarised things students said. Most also assisted students to complete tasks, reclarified the tasks and repeated or recorded what students needed to do. Very few provided the correct answer or deflected students’ questions, suggesting that these teachers fully engaged in dialogic classroom interactions and understood the value of feedback.

In a small number of lessons, teachers were observed providing students with opportunities to peer- and self-review work and prompting student self-reflection (AfL Strategies 4 and 5). In only five lessons were teachers observed asking their students to conduct a formal self-assessment (AfL Strategy 5), based on specific criteria, and in none of these were the teachers observed co-constructing the criteria with their students. However, it is important to note here that the observation schedule had limited focus upon student involvement in self-reflection and activating each other as resources for peer- and self-assessment (AfL Strategy 4). The analysis of the completed observation schedules did not reveal any obvious differences between the strategies used for different curriculum areas, although few teachers ventured into teaching lessons beyond literacy and numeracy.

Caution is advised in interpreting the findings outlined above. Firstly, due to the fact that two revisions of the observation sheet were made during the year of piloting, each iteration of the schedules was slightly different. This affected what could be gathered and compared. For example, the DEMFAP (original) had less room for observer comments and had fewer categories for ways in which students are activated as owners of their own learning. Nevertheless, in order to collect sufficient data from 10 teachers to enable analysis, all three versions of the observation sheet were included within the analysis. A further important limitation is that, although the teachers had some training in how to observe and record their observations, the focus for this activity was on evaluating and improving the observation schedules and providing peer feedback. Systematic training as observers would be necessary in order to use the observation schedule for research purposes. Beyond collating the information on the observation schedules and analysing the perception data from the teachers, we did not have the resources necessary to test the schedule with trained research observers and/or teachers skilled in the art of peer feedback. Increased resources would be necessary to rigorously test whether the observation schedule resulting from the pilot could be used validly for research purposes.

Research question 4: *What are teachers' perceptions about the efficacy and usefulness of the modified observation schedule for improving their assessment capability?*

Assessment capable teachers have the knowledge, skills, and volition to support students to become self-regulating, autonomous, life-long learners (Booth et al., 2016). To this end, data from the project were analysed to determine the nature of teachers' perceptions about the effectiveness and value of the modified observation schedule with reference to enhancing their AfL-related knowledge, understanding, and practice (skills). Teachers acknowledged that use of the schedule had encouraged them to "reflect on practice" (Tch 1, TS) and talked about how it served as a "guide to get you in that zone and thinking about what [you] are doing" (Tch 3, TS).

When asked for specific examples of how use of the modified schedule had helped improve AfL knowledge and practice, a number of teachers struggled to provide illustrations from experience. For some, use of the schedule and peer observation had failed to bring to the fore any substantive areas for improvement due to what they stated as their extensive experience as a teacher and/or the robust nature of their knowledge about and proficiency in AfL:

... to be honest I can't say I have used [the schedule] to inform my teaching ... I am an experienced teacher ... [my peer observer] thought I am doing enough. (Tch 2, RS)

[It] confirmed that I do do assessment for learning in my classroom ... it depends where the teacher is on their assessment for learning journey ... [for me] it confirms that [I am] doing the right things. (Tch 4, TS)

Others attributed failure to use the schedule to improve practice as due to what they saw as the impoverished state of their knowledge about AfL, such as "I don't know what it [AfL] is ..." (Tch 5, RS); and/or a mismatch between their understanding of AfL and the categories on the observation schedule. The latter caused one teacher to voice a lack of confidence in her understanding, "... to be totally honest, I [now] feel really insecure about my assessment for learning" (Tch 2, TS).

All teachers emphasised that users of the schedule "need to have ... a good understanding of what they are looking at to make this work" (Tch 1, TS). Teacher 1 at Totara School spoke at some length about how observers

need sufficient knowledge of AfL so they can identify the strategies they are seeing in colleagues' practice. This was considered a necessary condition for the provision of quality feedback and the identification of areas for improvement.

In Cycles 1 and 2 in particular, teachers also struggled to talk about specific ways in which they used the schedule to enhance their AfL capability as it seemed their efforts were directed at using the schedule to capture evidence in relation to specified categories and make suggestions about modifications to the schedule. Moreover, teachers' talk during peer debriefing sessions indicated they were preoccupied with noting and "ticking off" evidence of AfL practice rather than identifying areas for improvement:

I ticked I thought you helped children develop strategies ... I ticked 'no' for did the teacher collect something ... you had good questions ... (Tch 5 & 6, RS, PD Cycle 2)

This situation was exacerbated by there being no dedicated section on the earlier schedule/schedules for teachers to record areas for improvement and identify ways in which this improvement could be made:

but the sheet itself [the schedule] wasn't always clear, ... [nowhere to record] what the next step for him was on his assessment for learning journey ... (Tch 4, TS)

It was therefore not surprising to find teachers had some difficulty identifying and talking about ways in which the schedule had improved their AfL capability. Arguably, this finding reflects the nature of the project where the primary focus was on ascertaining how well the DEMFAP observation schedule and subsequent modifications captured New Zealand teachers' AfL practice, rather than its ability to develop teachers' assessment (AfL) capability.

Notwithstanding these issues, a close analysis of data revealed three aspects of AfL knowledge and practice, as represented in the modified observation schedule, that more than one teacher mentioned as an area they had identified for improvement or area where they had made changes to their practice due to the observation process. The first aspect dealt with the learning goals or learning intentions for the unit of work or lesson. Here, acknowledged features for development addressed a need to: provide students with a visual representation of the goal (Tch 1, TS); restate goals during lessons (Tch 6, RS) and; "clarify goals at all stages of the lesson and to check student understanding of the goals" (Eval Form, Tch 3, RS).

The second aspect spoke to student inclusion in the co-construction of success criteria. For example, one participant stated, "... so for [observed teacher] it was that ... [success criteria] hadn't been really deeply created with the children, [they] had just been created by the teacher" (Tch 4, TS referring to observation in Tch 2's, TS classroom). Likewise, another made the comment, "I need to spend more time encouraging students to co-construct their success criteria ..." (Eval Form, Tch 3, RS).

The third and final aspect was that of student self-reflection. Here, use of the modified schedule as part of the peer observation and feedback process stimulated, for some, a consideration of involving students in self-assessment/self-reflection. They made comments such as, "[I] will try to do more self-reflections with the children" (Tch 5, RS) and "I know it [the feedback session] made me more conscious of this part ... the self-reflection from students because I thought I am not actually doing that enough so I consciously decided I need to be doing more of that ... it was good pushing my practice ... I reflected on the fact there wasn't self-reflection [from the children], well none, but I could do that a lot better and more frequently ... [I could use] exit tickets ..." (Tch 1, RS). Teacher 1 at Rimu School was, however, the only person to talk about *how* she might go about enhancing her practice in relation to this aspect of AfL.

Virtually all of the teachers referred to one further aspect of AfL practice, questioning to elicit ideas from students. The ability to use different kinds of questions was noted by peer observers as a relatively strong feature of colleagues' AfL practice and, as such, thought it did not need improvement. Teacher 1 at Totara School was the only teacher who reflected on her ability to use questioning to elicit information from students. Interestingly, this reflection occurred, not as a result of feedback from her peer observer, but as a result of her observations in colleagues' classrooms:

[I have worked at increasing my practice in] eliciting information about what ... students know and can do. This has been an area of great personal growth. After observing teachers who are very good at eliciting information

from students I have been able to utilize some of their questions and discuss with teachers how they actually go about digging deeper to gain a better understanding of student knowledge ... The really sad thing is that up until this opportunity to observe I had felt I was eliciting information—it was only after the observations that I realized at what a surface level I had been operating". (Tch 1, TS)

In her final written evaluation, Teacher 1 at Totara School wrote at some length about how use of the modified observation schedule (including feedback from peers and observation in peers' classrooms) had catalysed personal reflection on and changes to aspects of her AfL practice (i.e., student access to learning goals; eliciting information through questioning; providing opportunities for students to act on feedback; examining current practice in student peer- and self-assessment). This teacher (along with her colleague Tch 2, TS) was relatively unique in terms of the breadth and depth of the way in which she used the observation schedule, peer observation, and feedback and self-reflection to enhance her assessment/AfL capability.

In summary, it appeared that, collectively, teachers' beliefs about the state of their AfL knowledge and skills, the project focus on determining the efficacy of the (modified) schedule to gather evidence of AfL practice and a determination to "tick off" items on the observation schedules, impacted on their ability to use the modified schedule to enhance their assessment capability. It was also noted that the modified version of the schedule should have a section where teachers could record areas for improvement or "where to next" and as a result of not having such a section, some teachers may have overlooked using the schedule for this purpose. Notwithstanding these issues, individual teachers found the modified schedule efficacious and useful in terms of promoting reflection and discussion on selected aspects of their AfL practice with some seriously engaging in this at a reasonably deep level.

4. Major implications for practice that derive from the findings

The purpose of this project was to pilot the DEMFAP observation schedules designed in the US in primary schools in New Zealand. It was clear that the DEMFAP observation schedule was not directly transferable to the New Zealand context. Rather, an observation instrument such as this needs to be adapted in language and form to reflect the local schooling context. Whilst the research team was able to adapt the schedule to make it better suit New Zealand primary schooling, particularly with respect to how teachers help students understand what quality looks like, further modifications appear to be required to amplify evidence about the ways in which they help students develop the metacognitive skills to evaluate their work and help students learn to use evaluative strategies and to modify their own work in line with the assessment capability framework (see Figure 1 above).

A second implication from this pilot project is that teachers need a very good understanding of what AfL is and what it means. This is required in order to recognise AfL strategies in action. They also need to know how to implement the five AfL strategies simultaneously in order to be able to use an observation schedule such as this in ways that further enhance/develop practice. In other words, they need to have a shared language and understanding about what constitutes AfL practice and this shared understanding also needs to be reflected in the observation schedule. Their understanding and use of AfL needs to be developed prior to using the schedule. This is because, if developed in tandem, the likelihood is that the use of the schedule will be prioritised over their understanding of what AfL is and how it is woven into classroom practice.

Thirdly, the productive use of an observation schedule such as the one piloted in this project will be dependent upon embedding its use within a strong professional learning programme where the purpose is focused upon enhancing the quality of teachers' AfL capability. From the findings of this project, it appeared that the use of the observation schedule was most effective as part of a peer observation and feedback process within a programme of professional learning. The teachers in this project cautioned against use of an observation schedule for appraisal or accountability purposes. Some had been observed previously as part of an AfL professional development programme. In comparison with the peer observation process used in this project, they stated

that a more formal type of observation by facilitators and/or senior management had not led to improved AfL practice, and for some, this had been an alienating experience. Thus, it appears that relational trust through peer observation and feedback was important for those who had made improvements to their AfL practice.

In conclusion, our project findings indicate that an observation schedule is one important aspect of a supportive professional development programme for enhancing AfL capability. The observation schedule needs to be consistent with the strategies for AfL used in that context. It was clear that, in the New Zealand context, building AfL capability needs to be a whole-school initiative conducted over an extended period of time incorporating both professional learning opportunities and feedback on practice through the use of an observation protocol. For this to happen, school leadership needs to be supportive of, and facilitate, AfL professional learning and development, enabling a supportive peer observation approach such as the one developed by Cycle 3 in this project.

References

- Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for Assessment in New Zealand (DANZ): Developing students' assessment capabilities*. Ministry of Education.
- Bell, N., & Cowie, B. (2001). *Formative assessment and science education*. Kluwer.
- Booth, B., Dixon, H., & Hill, M.F. (2016). Assessment capability for New Zealand teachers and students: Challenging but possible. *Set: Research Information for Teachers*, 2, 28–35.
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.
- Darr, C. (2018). A return to assessment for learning: Back to the future. *Set: Research Information for Teachers*, 1, 46–48.
- Education Review Office. (2015). *School evaluation indicators*. Author.
- Hill, M.F., Smith, L.F., Cowie, B., Gilmore, A., & Gunn, A. (2013). *Preparing primary and early childhood teacher education students to use assessment: Final summary report*. Retrieved from: http://www.tlri.org.nz/sites/default/files/projects/Hill_Final%20Summary%20Report_signed%20off.pdf
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109–138. <https://doi.org/10.1080/09585170600792712>
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 277–282. <https://doi.org/10.1080/09695940903319646>
- Ministry of Education. (2011). *Ministry of Education position paper: Assessment (schooling sector)*. <http://assessment.tki.org.nz/Media/Files/Ministry-of-Education-Position-Paper-Assessment-Schooling-Sector-2011>
- OECD/CERI (2008). *Learning in the 21st century: Research, innovation and policy*. Overview of the International Conference held in Paris, 15–16 May. <http://www.oecd.org/site/educeri21st/40820895.pdf>
- Parr, J. & Gadd, M. (2016). Generating positive outcomes by year 5–8 priority learners in writing: An inquiry into effective teacher practice. TLRI project in progress. <http://www.tlri.org.nz/tlri-research/research-progress/school-sector/generating-positive-outcomes-year-5-8-priority>
- Parr, J., & Hawe, E. (2009). *Measuring classroom practice in literacy*. <http://www.tlri.org.nz/tlri-research/research-completed/school-sector/measuring-classroom-literacy-practice>
- Poskitt, J., & Taylor, K. (2009). National education findings of Assess to Learn (AtoL). <https://www.educationcounts.govt.nz/publications/schooling/National-Education-Findings-of-Assess-to-Learn-AtoL-Report>
- Ruiz-Primo, M., Kroog, H., Richey, N., Iverson, L., Chzanowski, A., Shade, C., Silverstein, J., Zhao, X., & Sands, D. (2016, April). *Development, Evolution and adaptation of measures of formative practices: Lessons for the field*. Paper presented at the Annual Meeting of the American Education Research Association, Washington, D.C.
- Ruiz-Primo, M. (2017, April). *Variations of formative assessment practices across the instructional tasks in a lesson*. Paper presented at the Annual Meeting of the American Education Research Association, San Antonio, TX. <http://tinyurl.com/gv57v5o>
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433–449.
- William, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Learning Sciences International.

Research team

Associate Professor Mary Hill, The University of Auckland
mf.hill@auckland.ac.nz

Associate Professor Helen Dixon, The University of Auckland
Associate Professor Eleanor Hawe, The University of Auckland

Research partners

Beverley Booth and teacher researchers, Devonport Primary School
Jonathan Ramsay and teacher researchers, Edendale Primary School

Contact details:

Associate Professor Mary Hill
Learning, Development & Professional Practice, Education and Social Work, The University of Auckland,
DDI: + 64 9 623 8899, EXT: 48630, mf.hill@auckland.ac.nz