LEARNING PROGRESSIONS AND ONLINE FORMATIVE ASSESSMENT NATIONAL INITIATIVE

FINAL REPORT – ATTACHMENT 9

LITERATURE REVIEW: FORMATIVE ASSESSMENT EVIDENCE AND PRACTICE

EXECUTIVE SUMMARY

EXPERT PANEL

Dr Rod Lane (Chief Investigator) – Deputy Head of Department – Learning and Teaching

Professor Rauno Parrila (Chief Investigator) – Professor of Literacy (Reading)

A/Prof Matt Bower – Associate Professor of Technology-Enhanced Learning

Professor Rebecca Bull – Professor of Numeracy

A/Prof Michael Cavanagh – Associate Professor of Mathematics Education

Dr Anne Forbes – Senior Lecturer in STEM Education

A/Prof Tiffany Jones – Associate Professor in Sociology of Education

Dr Maryam Khosronejad – Education (Professional Development)

Dr David Leaper – Assessment Analysis and Development

Professor Liz Pellicano – Professor of Autism Education

Dr Sarah Powell – Creative Arts (Music and Dance)

Professor Mary Ryan – Professor of Literacy (Writing)

Dr Iliana Skrebneva – Senior Research Assistant

RECOMMENDED FULL CITATION

Lane, R., Parrila, R., Bower, M., Bull, R., Cavanagh, M., Forbes, A., Jones, T., Leaper, D., Khosronejad, M., Pellicano, L., Powell, S. and Ryan, M., Skrebneva, I. (2019) Formative Assessment Evidence and Practice Literature Review. AITSL: Melbourne

Contents

LIST OF ABBREVIATIONS AND/OR GLOSSARY

Term	Definition
Adaptive assessment	Assessment items and the nature of the feedback generated are based on the learner's current ability
AITSL	Australian Institute for Teaching and School Leadership
Assessment as learning	A type of formative assessment that focuses on teaching students the metacognitive processes to evaluate their own learning and make adjustments
Assessment for learning	A type of formative assessment that is used by teachers to gain an understanding of their students' knowledge and skills in order to guide instruction
ASSISTments	A web-based mathematics cognitive tutor. Like the adaptive systems described above, ASSISTments scaffolds problems into requisite skills and knowledge components.
Benchmark interim assessment	A comparison of student understanding or performance against a set of uniform standards within the same school year. It may contain hybrid elements of formative and summative assessments, or a summative test of a smaller section of the curriculum.
BYOD	Bring your own device
CAI	Computer-assisted interventions
CAT	Consequential Assessment Technique
CBM	Curriculum-based measurement
CCT	Classroom connectivity technology
CDDRE	Center for Data-Driven Reform in Education
CEM	Curriculum-embedded measures
COCA	Concepts of Comprehension Assessment
Cognitive conflict	A psychological state involving a discrepancy between cognitive structures and experience, or between various cognitive structures (i.e. mental representations that organise knowledge, beliefs, values, motives and needs). This discrepancy occurs when simultaneously active, mutually incompatible, representations compete for a single response.
Cognitive model	Model outlining the prerequisite cognitive and learning skills underlying successful progression – e.g. does the process require a significant amount of working memory, attention, motivation, persistency, cognitive ability, language skills, etc?
Conceptual change	Learning that involves the fundamental restructuring of students' pre-instructional ideas
Cronbach's alpha	Cronbach's alpha is a measure of internal consistency; that is, how closely related a set of items are as a group
Curriculum embedded measures	Formative assessments of recently taught content or skills designed to provide information that can guide instructional modifications for individual students
DAT	Diagnostic Assessment Tools
Data literacy	Data literacy is the ability to read, work with, analyse, and argue with data. Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data. Some of the competencies necessary to be able to demonstrate data literacy include: developing a habit of mind and practice

Term	Definition
	regarding data use; using inquiry processes; asking significant questions, collecting and organising data; knowing and understanding data properties; putting data in context (using pedagogical content knowledge); synthesising, probing and prioritising data; and transforming data into application.
Demonstrate	To show or make evident knowledge and/or understanding
DER	Digital Education Revolution
DFI	Dynamic Forecasting Intervention
Direct instruction	Teacher-centred pedagogy in which the teacher explicitly communicates a description of the concept to be learned or steps undertaken in a given practice to be learned
ERI	Early Reading Intervention
Experimental design	Experimental designs are used to examine the effect of a treatment or intervention on some outcome. In the simplest two-group case, a treatment is implemented with one group of participants (the treatment group) and not with another (the control group).
FAPD	Formative Assessment Professional Development
FAST-R	Formative Assessments of Student Thinking in Reading
Feedback	Informational response or information on reactions to an individual's, group's or organisation's performance (including the performance of a task or explanation of an idea, for example); intended as a basis for improvement.
Fidelity	The degree to which an intervention or program is delivered as intended
Formative assessment	Formative assessment is any interaction that generates information on student learning, which is then used by teachers and students to inform teaching and learning content and strategies (AITSL def: Evaluating student learning to provide feedback to students and devise/change teaching and learning programs – https://www.aitsl.edu.au/general/glossary).
Formative diagnostic assessment	A process of questioning, testing or demonstration, used to identify how a student is learning, where their strengths and weaknesses lie, and potential strategies to improve that learning. It focuses on individual growth.
I-Gmath	A synchronous peer tutoring system on mobile tablet devices
Interim assessments	Interim assessments are periodic diagnostic assessments, typically administered three or four times during the school year, to help teachers use evidence to differentiate instruction and make better instructional decisions, often in preparation for year-end summative assessments used as accountability measures.
ISI	Individualising Student Instruction intervention
ISI/A2i	A large-scale US software and instruction development project focused on improving K to 3 students' reading skills.
Learning progression	Learning progressions, also known as progress maps, developmental continuums and learning trajectories, describe research-based, descriptive continuums of how students develop and demonstrate deeper, broader, or more sophisticated understanding over time.

Term	Definition
Likert-type scales	A tool for measuring attitudes by asking people to respond to a series of statements about a topic, in terms of the extent to which they agree with them, and so tapping into the cognitive and affective components of attitudes.
MAP	Measures of Academic Progress
Mastery learning	Mastery learning model involves four components: defining mastery, planning for mastery, teaching for mastery, and grading for mastery. Formative assessments are used to provide both students and teachers with feedback about whether a particular instructional goal has been mastered. Students who do not meet the criteria for mastery are given correctives, such as alternative textbook readings, workbooks, or other varied learning tools. On completion of the correctives, the students take a second formative assessment. If they fail this test, they are given additional opportunities to study. Virtually all students achieve mastery before moving to the next unit.
mCLASS	A universal screener that measures the development of reading skills of all students in grades K–5 through two main assessments: Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Text Reading Comprehension (TRC) assessments.
Meta-analysis	A quantitative statistical analysis that is applied to separate but similar experiments of different and usually independent researchers and that involves pooling the data and using the pooled data to test the effectiveness of the results.
Metacognition	Metacognition is alertness to and reflection upon one's own or peoples' thought processes in learning and other contexts, or simply: thinking about thinking processes.
Metatalk	The use of language knowledge to have metalinguistic conversations
NAEP	National Assessment of Educational Progress
NAPLAN	National Assessment Program in Literacy and Numeracy
NCLBA	No Child Left Behind Act of 2001 (USA)
NPT	Nonreciprocal peer tutoring
OGT	Ohio Graduation Tests
PAL	Peer-assisted learning
PARS	Personalised Assessment Reporting System
Pedagogical content knowledge	Pedagogical content knowledge generally refers to teachers' expertise both in the specific subject being taught (the topics' information and skills) and in the best methods for teaching various components of the subject being taught (including different methods for the different information and skills in question, as needed).
PISA	Programme for International Student Assessment
PMA	Progress monitoring assessment, also known as learning progress assessment
POE	Predict, Observe, Explain
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
Progress monitoring	Progress monitoring is used to assess students' academic performance, to quantify a student rate of improvement or responsiveness to instruction, and to evaluate the effectiveness of instruction. Progress monitoring can be implemented with individual students or an entire class.

Term	Definition
PSADRI	Problem-solving Assessment, Diagnosis and Remedial Instruction
Reliability	Consistency or dependability of test performance across occasions, scorers and specific content
RPTMC	Reciprocal peer-tutoring-enhanced mathematical communication
Rubric	Rubrics are assessment tools with three characteristics: a list of criteria for assessing the important goals of the task; a scale for grading the different levels of achievement; and a description for each qualitative level.
Scripts	Scripts offer specific questions structured in steps to follow an expert model of approaching a task from beginning to end. They are designed to analyse the process being followed during a task, although they can also be used to analyse the final outcome.
SDA	Student Directed Assessment Process
Self-efficacy	A person's sense of being able to deal effectively with a particular task. Also, beliefs about personal competence in a particular situation.
SID	Scaffolding Instructional Discourse
Stratified randomised design	A probability sampling technique wherein the researcher divides the entire population into different subgroups or strata, then randomly selects the final subjects proportionally from the different strata
Summative assessment	Assessment of a test taker's knowledge and skills typically carried out at the completion of a program of learning, such as the end of an instructional unit
Systematic review	A systematic review is an appraisal and synthesis of primary research papers using a rigorous and clearly documented methodology in both the search strategy and the selection of studies
Task model	Model showing the sequence of activities to be successfully completed to meet learning outcomes and how learners typically progress through them (learning progression)
TTCT	Torrance Tests of Creative Thinking

1   Executive Summary

Formative assessment has been defined as any interaction that generates data on student learning and is used by teachers and students to inform teaching and learning, to address specific student learning difficulties and support learning growth over time.

A wide variety of assessment strategies, tools and resources currently exist to support and improve teachers' capacities to collect and analyse reliable data on student achievement and to help them adjust their teaching to meet each student's needs for enhanced learning outcomes. There is a widespread assumption in the academic literature that formative assessment leads to better learning outcomes for students.

Presently, we know little about the most effective ways of implementing formative assessment, including optimal school and educational system structures and supports. Systematic review of evidence is made more difficult by a lack of clarity and consensus regarding the nature and definition of formative assessment.

This review, commissioned by the Australian Institute for Teaching and School Leadership (AITSL), synthesises national and international research on the effective formative assessment practices of teachers and school leaders, including their current capacities, challenges and needs. It presents the findings of a review of peer-reviewed studies meeting robust experimental design criteria that examine formative assessment practices in Australian and international K–12 contexts. It further delivers an analysis and critical review of research relevant to formative assessment practices, including (but not limited to) the use of online assessment tools.

1.1   Key findings

1.1.1   Extent of Australian and international contemporary and seminal research on effective formative assessment practice (chapters 1–2)

A comprehensive database search of studies relevant to formative assessment practices in K–12 contexts identified 5867 studies. References were screened for quality and relevance (see Chapter 2). Further screening of papers against the tight inclusion criteria by five discipline experts reduced the pool of included papers to 71. Findings from these studies were then analysed to answer the review's focus questions. The authors note there are few rigorously designed experimental studies on formative assessment's impact on student learning, especially in Writing and the Arts.

1.1.2   Language, models and definitions of formative assessment (chapter 1)

The conception of formative assessment has broadened over the past 50 years, from the notion of formative evaluation (Scriven 1963) to a broad range of practices ranging from process-oriented 'formative learning assessment' to 'instrument-based formative assessment'. Unlike summative assessment, there is currently no agreed-upon definition of formative assessment, nor does formative assessment represent a well-defined set of practices. The varying definitions have made it difficult to compare studies of formative assessment's effectiveness.

This review describes two periods of development in the definition of formative assessment: (1) the 1960s to mid-2000s, when a loose consensus was developed and, (2) the period of conceptual confusion from 2000 onwards, when the boundaries of formative assessment have continued to expand.

Sparks' (2015) model of assessment types is presented as a potential solution to these definitional issues. Sparks differentiates formative learning assessment from formative diagnostic assessment, benchmark interim assessment and summative assessment. Chapter 1 concludes with an explanation of the formative assessment's domain general principles and an overview of the benefits of using technology to support formative assessment.

1.1.3   Findings for the impact of formative assessment on teaching practice and student learning progress/outcomes are mixed (chapters 3–7)

Research shows that formative assessment in different fields is often non-transferable and, when poorly designed, can lead to inaccurate information and ill-conceived pedagogical responses (Bennett 2011). The existing meta-analyses also have limitations that need to be considered. A number of these meta-analyses contain questionable claims about effect sizes that either summarise research too disparate to be synthesised meaningfully or are based on methods whose details were not published (Bennett 2011). To address this in the current review, we developed an evaluation framework for assessing the quality and rigour of formative assessment studies and meta-analyses and involved measurement and discipline experts in the evaluation and synthesis of this research.

Where studies report significant benefits, some isolate those benefits to specific groups of children (low achieving, high achieving). Others lack a sufficiently large sample size to look at effects for specific groups of students. Further, frequent omission of details regarding control group activities make determining formative assessment's impact on student learning more difficult. When studies involve a wholesale change to instruction (including changes to resources, professional development, time on task, etc.), it is impossible to isolate formative assessment's influence. Necessary information regarding fidelity of implementation, instructional/cognitive models and learning progressions is also missing from many studies.

Greater benefits do appear where targeted, individualised feedback is provided instantaneously and more frequently. Studies show a small number of yearly assessment points is not sufficient in providing timely feedback to students on specific learning tasks. Results suggest frequent and embedded formative feedback may be critical to effective implementation of formative assessment practices.

1.1.4   The effectiveness of particular tools and resources (including online tools) that support teachers' professional judgements of learners' needs and the implementation of formative assessment practices in school settings (chapters 3–7)

We found formative assessment tools/resources (online or otherwise) are most likely to be effective when based on:
1. a valid task model showing the sequence of activities to be successfully completed to meet learning outcomes and how learners typically progress through them (learning progression)
2. a valid cognitive model outlining the prerequisite cognitive and learning skills underlying successful progression (e.g. does the process require a significant amount of working memory, attention, motivation, persistency, cognitive ability, language skills, etc?).

Studies where formative assessments were followed by evidence-based interventions tended to produce better results.

Many of the reviewed studies included some technology/software component (e.g. mCLASS, ISI/A2i for reading) that is likely to be helpful in improving student learning outcomes. Studies with rigorous experimental designs and controls are, however, required to validate these causal links. Existing studies suggest there is a good reason to believe that experimental studies would show the positive effect of using technology for formative assessment.

Research showed a mixed picture regarding the use and effectiveness of online formative assessment interventions across different disciplines. The impact of technology-driven formative assessment on student learning outcomes depends on various factors, such as the characteristics of learners (low achieving, high achieving); study sample size; experimental design; and tool selection. The 'Using Sources Tool', for example, was found to be useful in facilitating teachers' evaluation of student writing tasks. Progress monitoring, in particular curriculum-based measurement (CBM) mazes, are shown to be effective in improving student reading. Adaptive computerised programs can be effective in improving learning in subjects where there are clearly identifiable skill hierarchies and relationships between skills.

Some outcomes, however, are difficult to assess using online tools, including 'creativity' in the Arts. A variety of assessment tools including observation; student–teacher collaboration; and self, peer and teacher feedback are required to effectively improve these skills.

The general principles for the effective application of formative assessment employment also apply to online assessment. Basing online assessments on a valid model of task components and the prerequisite cognitive and learning skills underlying successful progression is crucial. Interventions need to be evidence-based and aligned with validated learning progressions for the targeted concept or skill. We also found elaborate feedback with prompts is generally more effective than feedback that only recognises errors or provides correct answers.

Our research found optimal use of computer-based formative assessment is dependent on teachers' pedagogic preferences and orientations. When using technology to support formative assessment, it is vital teachers have the requisite hardware and software knowledge/skills and that formative assessment using digital technologies is supported and integrated within regular classroom activities. Teachers also need ongoing professional development to administer the assessments, interpret the results, make valid inferences and translate the information obtained to effective instruction.

1.1.5   Features of effective formative assessment professional development (chapter 8)

Attitudes to formative assessment can create barriers for implementation, especially when ideas and practices are incompatible with teachers' current views. Ongoing support is required if new ideas and strategies are to be effectively implemented. The research shows brief interventions, such as short-term, product-oriented workshops, are less likely to effectively change practice. By contrast, long-term, process-oriented professional development with ample opportunities for collaboration, feedback and discussion appears more effective in successfully changing teachers' classroom assessment practices. Professional learning that is work-embedded and situated within school needs is preferred over one-day workshops or formally presented interventions.

Several studies reported that practice-centred collaboration, often in the form of school-based professional learning communities, is a critical ingredient for effective formative assessment practices. Professional development is most effective when teachers actively engage in instructional inquiry as part of a collaborative professional community that is focused on instructional improvement and student achievement. Collectively sharing developmental work within the school site or across networks is also an important factor in formative assessment professional development success.

1.1.6   The optimal school and education system structures, supports and conditions for effective implementation of formative assessment practices, including implementation of assessment tools, and any barriers to their effective implementation (chapter 8)

We found that environmental conditions and teacher level factors both play a role in effective implementation of formative assessments. Research highlights the need for school leaders who understand formative assessment, can provide a rationale for its use and can create a supportive and non-threatening environment where the effective use of assessment data is modelled for staff. Leaders who can establish a school-wide formative assessment culture with vision and expectations for assessments, and a school climate promoting trust, mutual respect and cooperation will create the best environment for formative assessment success. This should further be reinforced with high-quality professional development and effective support for formative assessment implementation and a commitment to giving teachers regular, protected meeting times for meaningful examination of assessment practices.

It is important that accountability pressures on teachers do not lead to unintended impacts on instruction and assessment practices. Rather, decentralised organisational structures and distributed school leadership should focus on building a broader base of engagement and expertise, and a greater sense of shared vision and ownership. Strategically aligning expertise and resources to support teachers' learning about effective practice is necessary for achieving optimal implementation.

Teachers require both assessment knowledge and data literacy to effectively implement formative assessment. Increased focus on assessment literacy in initial teacher education and in-service teacher professional development is therefore needed. Promoting a classroom philosophy that regards mistakes as opportunities to learn and encourages honest reflection is key to achieving better results for students. Implementing effective formative assessment also requires sound pedagogical content knowledge, so that teachers can break down critical concepts, find appropriate entry points for all students, and redesign instruction to match students' assessed understandings and misconceptions.