

NSW Preschool Assessment Study: Review of formative assessment practices in early childhood settings



NSW Preschool Assessment Study: Review of formative assessment practices in early childhood settings

PREPARED FOR THE CENTRE FOR EDUCATIONAL STATISTICS AND
EVALUATION (CESE) AND THE EARLY CHILDHOOD EDUCATION
DIRECTORATE (ECED), NSW DEPARTMENT OF EDUCATION

Linda J. Harrison¹, Rebecca Bull¹, Sandie Wong¹, Sheena Elwick², and Belinda Davis¹, with the support of Veronika Kosourikhina¹, Nicky Spalding², Salima Yeung¹, Mandy Cooke¹ and Maggie Luck¹

¹Department of Educational Studies, Macquarie University

²School of Education, Charles Sturt University



Funded by NSW Department of Education (DOC 19/623423) 2019

Report prepared by Academy of Continuing Professional Development in Education, Macquarie University

TABLE OF CONTENTS

EXECUTIVE SUMMARY	3
GLOSSARY OF TERMS	9
LIST OF TABLES.....	10
LIST OF FIGURES	10
BACKGROUND	11
REQUIREMENTS.....	11
CONTEXT OF ASSESSMENT IN EARLY CHILDHOOD EDUCATION.....	11
THE CHALLENGE FOR EARLY CHILDHOOD EDUCATION	13
RESEARCH AIM AND APPROACH.....	14
LITERATURE REVIEW.....	16
METHODOLOGY	16
SUMMARY OF FINDINGS	20
DOMAIN GENERAL TOOLS.....	20
1. TEACHING STRATEGIES GOLD	20
2. CHILD OBSERVATION RECORD (COR).....	24
3. DESIRED RESULTS DEVELOPMENTAL PROFILE (DRDP)	26
DOMAIN SPECIFIC TOOLS	33
2. INDIVIDUAL GROWTH AND DEVELOPMENT INDICATORS – EARLY LITERACY (EL-IGDIS)....	35
PROFESSIONAL LEARNING AND DEVELOPMENT.....	36
METHODOLOGY	38
SUMMARY OF FINDINGS	40
DIVERSE TYPES OF FORMATIVE ASSESSMENT TOOLS USED IN DAY-TO-DAY PRACTICE	41
PRACTICE MODELS FOR INTEGRATING EFFECTIVE FORMATIVE ASSESSMENT TOOLS	60
EVALUATION OF THE FA TOOLS BEING USED IN THE SERVICES	62
VALIDITY OF FA TOOLS	62
MODELS OF PROFESSIONAL DEVELOPMENT ACCESSED BY SERVICES TO SUPPORT FA	64
SUPPORTS AND BARRIERS TO USING FA TOOLS	66
CURRENT PRACTICES, CONCERNS AND CHALLENGES	67
CLARIFICATION OF THE PURPOSE OF ASSESSMENT.....	68
MEETING A NEED	68
SELECTING / ADAPTING A FORMATIVE ASSESSMENT TOOL OR TOOLS	69
PROFESSIONAL LEARNING.....	70
POLICY ISSUES IN THE IMPLEMENTATION OF FORMATIVE ASSESSMENT	71
REFERENCES	72
APPENDIX A: LITERATURE SEARCH KEYWORDS	77
APPENDIX B: LIST OF FORMATIVE ASSESSMENT TOOLS	78
APPENDIX C: DETAILED DESCRIPTIONS OF FORMATIVE ASSESSMENT TOOLS AND SUMMARIES OF RESEARCH EVIDENCE	81

APPENDIX D: PROFESSIONAL LEARNING SEARCH RESULTS	111
APPENDIX E: ADDITIONAL RATINGS FOR EACH TYPE OF ASSESSMENT TOOL	116

EXECUTIVE SUMMARY

BACKGROUND

This research was commissioned by the New South Wales (NSW) Department of Education (DoE) to investigate formative assessment tools for early childhood education (ECE) services in NSW, focusing on children in their year before school. The purpose of the research project, described by five areas of investigation, was to:

- assess the validity of different formative assessment tools and any evidence that their application has led to improved teaching practice and child development outcomes;
- identify the extent to which any of the effective tools have been currently employed in the NSW ECE sector;
- indicate the extent to which the observed outcomes of effective formative assessment tools align with the Early Years Learning Framework;
- indicate whether the tools have been demonstrated to be effective for students with additional needs;
- Consider effective tools that can be embedded in the NSW ECE sector, including by providing information on best-practice models of delivery of professional learning and effective tools for the sector.

The scope of the study comprised two complementary components: a review of the literature on formative assessment tools used in early childhood education and care settings for children aged 3- to 5-years; and stakeholder consultations with ECE service providers and staff who are directly involved in the selection, use, and interpretation of formative assessment tools and practices.

The investigation took account of the current context of assessment requirements and practice in ECE services in NSW, which is underpinned and informed by three policy documents:

- the Early Years Learning Framework for Australia (EYLF);
- the National Quality Standards (NQS) particularly Standard 1.3, Assessment and Planning and Standard 7.2, Leadership; and
- Australian Professional Standards for Teachers (Proficient Level).

Teachers, educational leaders, and educators working in ECE services are accountable for recording, reflecting on, and using information about how children are developing and progressing their interests, and what they are learning. However, whilst there is an expectation that this information is used to plan integrated and meaningful opportunities and experiences for children's learning, there are no guidelines or recommendations about what assessment tools should be used.

STUDY DESIGN

The review of literature and stakeholder consultations were undertaken simultaneously. For each component, a set of specific research questions were developed in consultation with the DoE to operationalise the five research areas.

Literature Review

The literature review involved a structured process designed to gather all relevant studies on formative assessment in early childhood education through top-down and bottom-up literature searches. Top-down searches included: formative assessment tools and evidence for their effectiveness in Australia and globally, professional development specific to implementation of formative assessment, and supports and barriers to implementation of formative assessment. Bottom-up searches included looking for evidence supporting the

use of specific tools that were either known to be used, or could be viably used, by educators in ECE services, for formative assessment. A rigorous screening process was undertaken to evaluate the studies for inclusion in the literature review.

Studies that passed the screening process were reviewed in full by a member of the research team. This process identified 22 assessment tools, which were then evaluated against an Evaluation Framework developed by the research team to determine:

- Study design and quality;
- Study sample characteristics (child ages, inclusion of special populations, ECE setting type);
- How the tool was administered (by whom, time to complete, frequency of completion);
- Feedback (what information does the tool provide, how is this used by teachers);
- Content (developmental domains that are linked to the EYLF Learning Outcomes);
- Psychometric information (validity, internal reliability, scale structure);
- Professional development (duration of the program, who attended, resources);
- Outcomes (for children and teachers).

The literature review on professional development for formative assessment was supplemented by a search of ECE providers of professional learning in NSW and Australia to ascertain what was available to educators.

Stakeholder Consultations

A case study approach was used to gather data in eight selected ECE services for children aged 3- to 5-years that had shown exemplary and innovative formative assessment practices, as identified by the NQS Assessment & Rating process. A diversity of services was selected from an initial set of services identified by the DoE, ensuring representation of four characteristics:

- location (metropolitan, regional, remote);
- type of program (preschool; long day care);
- type of Approved Provider organisational structure (for-profit; not-for-profit);
- size of Approved Provider organisation (standalone service; 2-7 services; > 7 services).

A 2-day visit to each service was made by a member of the research team. Interviews were conducted with a total of 35 participants, including eight Centre Directors (CD), seven Educational Leaders (EL), eight Early Childhood Teachers (ECT), nine Educators (Ed) including specialist support staff / liaison officers (SSLO), and three representatives of Approved Provider (AP) organisations.

Each participant was interviewed individually and asked to bring de-identified examples of assessment and documentation to illustrate practices. Interview questions covered a range of topics related to the use of formative assessment tools, including ratings for specified assessment tools, and open-ended comments and explanatory text. Data analysis used descriptive statistics and qualitative coding methods.

KEY FINDINGS

Findings are summarised under each of the five research areas, drawing on both components of the investigation: the review of literature and professional learning, and the stakeholder consultations

Assessment of the validity of different formative assessment tools and evidence that their use has led to improved teaching practice and child development outcomes

A critical review of assessment tools identified in the literature review confirmed widespread use in ECE services in the United States (US) of formative assessment tools whose content covered a number of different domains of learning and development; that is, they were Domain General tools. Three of these were selected for detailed review:

- *Teaching Strategies Gold* (TS GOLD), which is associated with the widely used Creative Curriculum for Preschool (<https://teachingstrategies.com/solutions/teach/preschool/>);
- *Child Observation Record* (COR), which is linked to the Highscope Curriculum (<https://highscope.org/our-practice/preschool-curriculum/>);

- *Desired Results Developmental Profile (DRDP)* (www.desiredresults.us), which was developed by the California Department of Education to support educators across a wide range of curriculum models, including Head Start.

Each of these three tools had been shown to meet criteria for internal and inter-rater reliability, and for construct validity when compared with standardised tests of language, literacy and numeracy, as well as social functioning. Further, TS GOLD and the DRDP have been shown to provide valid ratings of developmental progress across a range of domains.

Evidence that the use of these assessment tools resulted in a change in teacher practice was mixed. In one study it was found that teachers using TS GOLD prioritised certain play types or experiences that yielded information related to TS GOLD objectives, and were less likely to rely on naturally occurring, spontaneous opportunities for observing children's learning. On the other hand, a large study of TS GOLD found that most teachers used assessment results to help them individualize instruction, and to identify individual learning needs.

In regard to evidence that using formative assessment impacts on child outcomes, reviewed studies for TS GOLD and COR were part of broader evaluations of curriculum implementation in which the tool is just a small part of the package used for individualised programming. It was impossible, therefore, to evaluate the direct contribution to improved child outcomes of using the tool. The DRDP, on the other hand, is not aligned with a specified curriculum. DRDP results can be processed and aggregated as a 2-page individual rating record that summarizes the child's progress. Teachers report that a strength of the DRDP is its usefulness for gauging students' progress and organising levelled groups according to children's respective strengths and needs. At the time of the review, however, no studies had tracked the impact using the DRDP on child outcomes.

In relation to the validity of Learning Stories and other observational methods of assessment used in the NSW case study services, there was no evidence that standard measures of reliability and validity were met. In fact, the developer of learning stories, Margaret Carr, has suggested that the traditional notion of validity is inappropriate for learning stories. For this reason, these approaches do not lend themselves to an evaluation of changes in children's learning over time, which makes it difficult to assess their impact on child outcomes.

It was clear, however, from the stakeholder consultations, that the use of formative assessment does impact teaching practice, particularly in the sense that it is seen as an important part of meeting NQS and EYLF requirements. The use of formative assessment was linked to the planning cycle of observe, critically reflect and interpret, plan and provide experiences, observe and so on. Participants described using different types of tools, such as observations, learning stories and reflections, together as an informative planning process. Assessment prompted critical reflection on practice and informed the choice of teaching strategies to extend children's learning. However, the tools that were used in the NSW case studies did not provide definitive evidence that their use had an impact on children's developmental outcomes,

Identification of the extent to which any of the effective tools are currently employed in the NSW ECE sector

Formative assessment tools described by CDs, ELs, ECTs, Eds and SSLOs in the stakeholder consultations included written observations with or without photographs, learning stories, child portfolios, floor books, photographs alone, checklists which included developmental, skills-based, and school readiness items, child-completed assessment tools, family-completed assessment tools, on-line assessment tools, service-designed tools, and externally sourced tools. Participants also mentioned their use of digital apps, daily journals, individualised learning plans, transition of room assessment and collections of children's work samples. Teacher reflections also featured in the broad suite of assessment methods.

Overall, the tools that were rated as most useful (ratings over 4.5 out of 5) by at least half of the participants were: written observations, learning stories, service-designed and family-completed tools. A substantial number of services were using an externally designed, on-line or their own designed checklist as a key part of the formative assessment. The use of a tool where a 'rating' has to be recorded across all learning and development areas forces educators to look holistically at the child.

Indication of the extent to which the observed outcomes of effective formative assessment tools align with the five EYLF Learning Outcomes

Assessment tools identified in the literature review were evaluated on the extent to which the domains or specific items covered by the tool aligned with the five EYLF Learning Outcomes. TS GOLD included assessment items that aligned with EYLF Learning Outcome (LO) 3 (Children have a strong sense of wellbeing), LO4 (Children are confident and involved learners), and LO5 (Children are effective communicators). COR and DRDP included items that aligned with all five EYLF Learning Outcomes, including LO1 (Children have a strong sense of identity), and LO2 (Children are connected with and contribute to their world).

Learning Stories, due to their subjective nature, could potentially address all of the EYLF outcomes, although alignment is difficult to map given that there are no objective skills or outcomes stated. Some authors expressed concern that the focus on learning dispositions may lead to neglect of the development of knowledge and skills. This also suggests a less than comprehensive alignment with the EYLF LOs.

The interviews with CDs, ELs, ECTs, Eds and SSLOs overwhelmingly endorsed the view that all of the assessment tools that were currently being used were suitable for all five EYLF LOs. Very few examples were given of selecting a specific tool for assessing a specific LO.

Have the tools been demonstrated to be effective for students with additional needs?

The assessment tools identified in the literature review were evaluated on their suitability for diverse populations, including children who were developmentally delayed, culturally and linguistically diverse, and from low-income families. TS GOLD is inclusive of English language learners (ELLs) and children with disabilities as well as typically developing children and those who demonstrate competencies beyond developmental expectations. COR is designed to capture the developmental trajectories of all children. The development of the DRDP followed the principles of Universal Design with the goal of ensuring that all children have the opportunity to demonstrate their knowledge and skills; it gives specific consideration for children who are dual language learners. DRDP resource material provides categories of adaptations to support specific groups.

Learning Stories should be suitable for any population, although studies suggest that educator's biases could lead to unfavourable interpretations of events and learning for some children, and the lack of a systematic, comprehensive approach may lead to overlooked development in some areas for any child.

Most of the tools described by CDs, ELs, ECTs, Eds and SSLOs in the stakeholder consultations were felt to be suitable for children with a range of abilities (ratings over 4.5 out of 5). Participants were less positive about the suitability of these tools for children who do not have English as their home language (ratings of 3.5 to 4.5), suggesting a degree of concern over the effectiveness of the tools for students coming from a language background other than English.

Best-practice models of delivery of professional learning for embedded effective formative assessment tools in the NSW ECE sector

Professional learning, development and training is an essential component of effective assessment practice in ECE. Teachers' ability to reliably and accurately score standardised assessment tools, such as TS GOLD, COR and DRDP, rely on teachers' ability to observe children's naturally occurring interactions and to make judgments about what children can and cannot do. Professional development is included for TS GOLD (compulsory 2 days initial training plus booster training), but less evident for COR (training offered) and the DRDP (training and resources available but uptake of formal training is variable). Studies noted that teachers are eager to learn more about the assessment tool they are using and want more time for professional development activities. Ongoing professional development relies on informal support from supervisors on interpretation, goal setting and planning.

The wider review of research suggested that not all ECE teachers are skilled observers, and some have difficulty assigning scores in a systematic and defensible manner. Therefore, if a standardised tool is to be adopted by ECE services, it is essential that training material and supports are available and rigorously applied.

Our search for professional learning opportunities in NSW and Australian revealed that little is available on the topic of assessment in ECE and that, of those identified, none directly used the term assessment in the title. The search and review process showed a clear need for specific professional opportunities on formative and summative assessment in ECE across Australia. Through strengthening understandings of what these types of assessment looks like, educators may be better able to examine, adapt and articulate formative and summative assessment within their workplaces

The research literature on models of effective professional learning in ECE underlines the importance of organisational support for pedagogical leadership and mentoring. This approach was endorsed by CDs, ELs, ECTs, Eds and SSLOs, who reported that in-house development and support from staff within the larger organisation was the most common form of professional development. This included induction, mentoring, and support from colleagues. Participants said that the support and leadership of ELs who know what to do was key to the effective use of formative assessment tools and processes.

OVERVIEW OF FINDINGS AND FURTHER RESEARCH AREAS

Current Practices, Concerns and Challenges

This research investigation commissioned by CESE and the ECED has clearly tapped into an issue of growing concern and a challenge that the ECE field as a whole will need to face. Addressing the challenge of adopting effective assessment practices in ways that will be acceptable and appreciated by the field will need to draw on principles of Implementation Science. Redding et al. (2017) describe implementation science as the study of processes and conditions that promote or impede the effective take-up of evidence-based practices in real-world contexts. Implementation success relies on educator buy-in and participation, and a supportive organisational context.

It is clear that services are looking for or designing their own formative assessment tools. In moving forward to support this direction in ECE, the selection of tools should aim to meet criteria for psychometric validity as well as for social validity (Bagnato, Neisworth, & Pretti-Frontczak, 2014). Social validity refers to the tool's ease of use, its accessibility, its acceptability to teachers and families, and its suitability for supporting communication and collaboration across key stakeholders – ECE educators, schools and teachers in the first year of school, and families.

Clarification of the Purpose of Assessment

The findings highlight the tension between pedagogical documentation for the purposes of planning for individual and group learning, and documentation for tracking the process and progression of children's learning, that is, for formative assessment.

It will be important for ECE providers, policy makers and researchers to work together to clarify the purpose(s) of assessment the year before school, and particularly in relation to gathering and providing information that supports children's transition to school.

Meeting a Need

ECE services are looking for or developing their own instruments to record change over time, but there are no universally advised formats in NSW and no evidence that the formats services are using have psychometric validity. There is no standardised measure of change or evidence that change is related to the educational program.

Further research is needed to produce or adapt a 'standard' developmental checklist that can be used by NSW services to identify the individual competencies and learning needs of each child, and track learning progression over time. The skills, attitudes and understandings assessed by an adapted or locally-designed tool should directly map onto NQS standards for assessment in ECE and also link to indicators of development and learning identified in the EYLF. Once available, a reliable and valid tool could help teachers monitor individual children's progress and potentially also be used to create a classroom profile of individual differences in children's content knowledge and ability levels to inform planning for teaching and learning.

Selecting / Adapting a Formative Assessment Tool or Tools

There is no currently available tool that can be implemented in ECE services in NSW. Further research could be undertaken by forming a collaborative working group to conduct a thorough review and evaluation of a standardised, psychometrically valid tool such as the Desired Record of Developmental Progression (DRDP) which was developed by a consortium of experts for use across a variety of services and curricula.

A co-design process with researchers, practitioners, and the tool-developers would evaluate the tool, for its:

- alignment with EYLF learning outcomes in the year before school;
- alignment with expectations of school readiness as children transition to kindergarten;
- social validity, acceptability and relevance to the everyday work of ECE teachers;
- alignment with current practices for observing, recording and collecting evidence of children's learning experiences and interests in ECE services;
- psychometric reliability and construct validity when used with Australian children;
- impact on practice in the classroom;
- impact on outcomes for children.

Professional Learning

Given the limited focus on assessment in ECE training and professional learning, we have observed the following needs; that:

- pre-service and in-service training include evidence related to the implementation of different formative assessment tools to allow practitioners to think critically and creatively;
- ongoing professional development is required to support all aspects of formative assessment;
- Educational Leaders and Centre Directors receive specialist professional development to understand the expectations for leadership when planning and implementing formative assessment that meets requirements and considers the culture of the service.

The evaluation of professional learning for assessment is also needed, particularly in order to track changes or improvements to teaching strategies.

Policy issues in the Implementation of Formative Assessment

A consistent theme, in both the Stakeholder Consultation interviews and the Literature Review, was educators' concern about the time-consuming nature of recording assessments. Further research is needed to investigate models of effective assessment practice that consider efficiencies in resourcing so that teachers are able to plan diverse activities and compile records on individual children. Staffing models that involve a specialist support worker or liaison officer who can facilitate engagement with children in smaller groupings and more individualised planning are also worthy of further research.

GLOSSARY OF TERMS

AP	Approved Provider
CD	Centre Director
CELA	Community Early Learning Australia
DoE	Department of Education
ECA	Early Childhood Australia
ECE	Early Childhood Education
ECT	Early Childhood Teacher
Ed	Educator
EL	Educational Leader
EYLF	Early Years Learning Framework
FA	Formative Assessment
NQS	National Quality Standard
QA	Quality Area
SSLO	Specialist Support / Liaison Officer

LIST OF TABLES

Table 1: Matching of Five Specified Research Areas to Research Questions Driving the Study Methodology..	15
Table 2: Responses to Use (yes/no) and Perceived Usefulness (1 – 5 rating) of Assessment Tools	41

LIST OF FIGURES

Figure 1: Example of Observation Record – Group of Children.....	43
Figure 2: Example of Observation Record – Individual Child	43
Figure 3: Example of Learning Story	46
Figure 4: Example of Floor Book.....	48
Figure 5a: Example of Service-Designed Checklist	50
Figure 5b: Jottings.....	51
Figure 5c: Educational Summary	51
Figure 6: Example of School Readiness Checklist.....	52
Figure 7: Example of Child-completed Assessment	54
Figure 8: Example of Child-completed Assessment as part of an educator-completed checklist	54
Figure 9: Example of Family-completed Assessment	56
Figure 10: Example of Family-completed Assessment Tool.....	57
Figure 11: Daily Journal - shared on Digital App.....	58
Figure 12: Example of Teacher Reflection on a Learning Experience.....	60

BACKGROUND

REQUIREMENTS

This research was commissioned by the New South Wales (NSW) Department of Education (DoE) to investigate formative assessment tools for early childhood education (ECE) services in NSW, focusing on children in their year before school. Formative assessments are designed for educators to gather evidence of children's learning in order to clarify the educator's understanding of the child's capabilities and to inform individualised teaching. The use of formative assessments can encourage educators to be more systematic and consistent in the way they consider each child in all areas of learning and development.

The overall intended purpose of the research project was to identify best practice formative assessment tools for children in the year before school and evaluate the suitability of these tools for early childhood education (ECE) services in NSW. The objectives, identified in the Request for Quote (RFQ) were to: provide evidence of effectiveness of the use of formative assessment tools against measures of teaching quality and the outcomes of the Early Years Learning Framework EYLF (2009); identify and consider best practice models for integrating effective formative assessment tools in early childhood setting(s) and provide an assessment of the feasibility of such models in the NSW ECE sector; and assess the risks and costs of different tools on services, teachers and children.

The scope of the work comprised five research areas:

- i. assess the validity of different formative assessment tools and any evidence that their application has led to improved teaching practice and child development outcomes;
- ii. identify the extent to which any of the effective tools have been currently employed in the NSW ECE sector;
- iii. indicate the extent to which the observed outcomes of effective formative assessment tools align with the EYLF;
- iv. indicate whether the tools have been demonstrated to be effective for students with additional needs;
- v. consider effective tools that can be embedded in the NSW ECE sector, including by providing information on best-practice models of delivery of professional learning and effective tools for the sector.

The recommended methodology identified in the RFQ was a literature review and engagement with between 2-3 ECE services from remote, rural and metropolitan regions of NSW.

CONTEXT OF ASSESSMENT IN EARLY CHILDHOOD EDUCATION

EARLY YEARS LEARNING FRAMEWORK

Assessment for learning and development is defined in the Early Years Learning Framework for Australia (EYLF) as:

...the process of gathering and analysing information as evidence about what children know, can do, and understand. It is part of an ongoing cycle that includes planning, documenting and evaluating children's learning (Australian Government Department of Education, Employment & Workplace Relations [DEEWR], 2009, p. 17).

The forms that this process should take are not prescribed in the EYLF. Educators are expected to assess children's learning in ways that are authentic and collaborative, and that highlight children's strengths. The expectation is that educators will "search for appropriate ways" and use a "variety of strategies" and methods to "capture and validate the different pathways that children take toward achieving" the five Learning Outcomes set out in the EYLF, as well as make "the process of learning visible" (p. 17). By viewing assessment as "making learning visible", the EYLF endorses the benefits of assessment as part of delivering a high-quality early childhood program (Giuduci, Rinaldi & Krechevsky, 2001).

Whilst not using the term 'formative assessment', the intent of the EYLF aligns with the accepted, broad definition of formative assessment as a "process that teachers employ to collect and use assessment

information to tailor instruction to the individual needs of children” (Riley-Ayers, 2014, p. 4). This process is explained for teachers in more detail, along with a visual diagram, as the Early Years Planning Cycle – Belonging, Being and Becoming in the Educators’ Guide to the EYLF (DEEWR, 2010). The Guide introduces educators to the term assessment tools, with the recommendation that educators “use a range of assessment tools and methods” (p. 38). The Guide sets out four principles and criteria to guide selection of tools.

Assessment practice:

Ethical assessment... provide children with opportunities to confidently demonstrate their capabilities

Dynamic assessment... in the context of meaningful, supportive and respectful interactions

Forward thinking assessment ... ability to assess children’s potential, rather than just their actual development/learning

Child oriented assessment ... methods and tools where children can assess themselves

Assessment content:

Assessment should match the curriculum ... content is appropriate and assesses important skills or understandings that are rich, complex and integrated; methods are appropriate and not too time-consuming

(extract from DEEWR, 2010, p. 39, citing Fler, 2008)

Notably, these guiding principles do not extend to the more specific expectations of formative assessment that the tools “must be reliable and valid” (Riley-Ayers, 2014, p. 5.). This is a critical criterion, which ensures that the evidence, or data, provided to teachers can inform individual teaching and learning experiences, and track children’s progress towards learning goals.

NATIONAL QUALITY STANDARD

Assessment is a key feature of Australia’s National Quality Standard (NQS) (ACECQA, 2018) for ensuring the achievement of quality ECE by service providers. Assessment features in Quality Areas 1 and 7.

Within Quality Area 1, Educational Program and Practice, is Standard 1.3, Assessment and Planning with the expectation that “educators and co-ordinators take a planned and reflective approach to implementing the program for each child” (ACECQA, 2018, p. 15). This is achieved by three Elements:

1.3.1 Assessment and planning cycle: Each child’s learning and development is assessed or evaluated as part of an ongoing cycle of observation, analysing learning, documentation, planning, implementation and reflection. Element 1.3.1 is underpinned by National Regulation 74 Documenting of child assessments or evaluations for delivery of educational program, and in NSW by Regulation 274A Programs for children over preschool age.

1.3.2 Critical reflection: Critical reflection on children’s learning and development, both as individuals and in groups, drives program planning and implementation.

1.3.3 Information for families: Families are informed about the program and their child’s progress.

Within Quality Area 7, Governance and Leadership, is Standard 7.2, Leadership, which states: “Effective leadership builds and promotes a positive organisational culture and professional learning community” (ACECQA, 2018, p. 58). Standard 7.2 is supported by three Elements:

7.2.1 Continuous improvement: There is an effective self-assessment and quality improvement process in place.

7.2.2 Educational leadership: The educational leader is supported and leads the development and implementation of the educational program and assessment and planning cycle.

7.2.3 Development of Professionals: Educators, co-ordinators and staff members’ performance is regularly evaluated, and individual plans are in place to support learning and development.

The effective implementation of Quality Area 1, Standard 1.3 is closely linked to the services' support for the role of the educational leader, and for professional development for educators.

Data for all ECE services' ratings on the National Quality Standards are held by the Australian Children's Education and Care Quality Authority (ACECQA). A recent analysis of ACECQA data for NSW, conducted by CESE for the DoE, showed that services that performed well on Elements 1.3.1 and 7.2.2 were most prevalent in services with an overall rating of Exceeding NQS. A recent report from ACECQA (2019) notes that 28% of long day care services and 60% of preschools have an overall rating of Exceeding NQS.

NSW PROFESSIONAL STANDARDS FOR TEACHERS

National and State Regulations stipulate requirements for ECE to be staffed by one or more qualified early childhood teachers (ECTs). Accreditation of teachers, including ECTs, is underpinned by professional standards. The importance of assessment is underlined Australian Professional Standards for Teachers at Proficient Teacher, which include:

- selecting and using informal and formal assessment strategies that are differentiated for the specific learning needs of students across the full range of abilities;
- interpreting and using student assessment data to evaluate student learning and modify teaching practice; and
- diagnosing barriers to learning and challenging students to improve their performance (NESA, 2018).

Teachers are also encouraged to complete a summative assessment tool, the Transition to School (T2S) Statement, which provides a summary of "the child's strengths, perspective, and personality" that can be shared with families and the kindergarten teacher at school entry (NSW Dept of Education, 2018, p. 1). The T2S Statement should draw on the child's progress towards the five EYLF Learning Outcomes and align with the Early Stage 1 syllabus.

THE CHALLENGE FOR EARLY CHILDHOOD EDUCATION

Teachers, educational leaders, and educators working in ECE services are accountable for recording, reflecting on, and using information about how children are developing and progressing their interests, and what they are learning. While the EYLF (DEEWR, 2009) and The Educators' Guide (DEEWR, 2010) direct educators to use assessment to look at individual pathways for children, liaise with families and plan integrated and meaningful opportunities and experiences for children's learning, no specific assessment tools are recommended.

The EYLF recognises that how assessment will look in ECE settings will likely be just as diverse as the communities it is being used in across Australia. Some ECE settings work with a specific assessment tool they are required to use by their overarching governing organisation, whereas others develop their own approaches (Patterson & Fleet, 2011).

For formative assessment tools to be used successfully and effectively, educators working in ECE settings need professional development and training in the implementation of specific tools, and the use of assessment data to judge children's progress and improve their teaching practices (Riley-Ayers, 2014). To meet this need, the Victorian Curriculum and Assessment Authority (VCCA) commissioned a series of reviews of assessment tools for three of the five EYLF Learning Outcomes: Children have a strong sense of wellbeing (Marbina et al., 2015); Children are confident and involved learners (Cloney Jackson & Mitchell, 2019); Children are effective communicators (Verdon et al., 2018). Each review identified a list of existing instruments that ECE educators might use; however, the effectiveness of these as a formative assessment tools was not tested.

WHAT IS CURRENTLY RECOMMENDED FOR ECE EDUCATORS?

Cloney et al. (2019, p. 7) note that while “assessment of young children’s learning is a contested issue in some areas of ECE practice”, with concerns about methods from school education ‘pushing down’ into ECE services, there is “a place for rigorous assessment” of “the constructs of learning that are unique to the early years” as set out in the EYLF and the Victorian Early Years Learning and Development Framework (VEYLDF) (Department of Education and Training [DET], 2016).

While it is important to not dismiss concerns that standardised assessment tools may not recognise that young children can demonstrate learning, skills and abilities in a variety of ways, and may not be sensitive to children’s cultural and linguistic diversity (Bowman, Donovan & Burns, 2001), appropriate assessment tools that meet specific criteria for reliability and validity can provide evidence of learning and learning progression across the preschool years.

Implementing such an approach in ECE settings would align with the primary goal identified in the Australian Government’s *Report of the Review to Achieve Educational Excellence in Australian Schools* (Gonski, et al., 2018).

Priority one: Deliver at least one year’s growth in learning for every student every year, where “student growth is a measure of the individual progress a student makes over time along a defined learning progression” (p. 5).

The search for an appropriate tool to measure learning progress in the years before school was the impetus for work conducted by Early Start, The University of Wollongong, which developed an assessment instrument, the Early Years Toolbox (Howard & Melhuish, 2017), to assess children’s language, numeracy, and self-regulation and social development skills using brief game-like methods suitable for delivery on an iPad (<http://www.eytoolbox.com.au/>). A pilot study is currently underway to train early childhood staff to use the Toolbox as a formative assessment tool with 3 to 4-year-old children, and to evaluate its impact on professional practice. Results of the pilot are due to be released in 2020 (<https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/early-years-toolbox/>).

RESEARCH AIM AND APPROACH

In sum, recent work in Australia concludes that although many tools have been developed and most have undergone some form of validation, few are available and applicable for use by early childhood professionals in their day-to-day practice:

Without the evidence that assessment provides, early childhood professionals may struggle to know whether their pedagogy is having an impact, or where gaps remain in children’s knowledge and skills. (Cloney, et al., 2019, p. 31).

The approach developed by an expert research team led by Macquarie University with colleagues at Charles Sturt University to address this urgent issue comprised two parallel components:

- a review of the literature on formative assessment used in early childhood education and care settings for children aged 3- to 5-years; and
- stakeholder consultations with ECE service providers and staff who are directly involved in the selection, use, and interpretation of formative assessment tools and practices.

Each component was developed in consultation with the DoE. For each, a set of specific research questions were identified for each component of the study in order to operationalise the five research areas identified in the RFQ (see Table 1).

Table 1: Matching of Five Specified Research Areas to Research Questions Driving the Study Methodology

Research Areas in RFQ	Literature Review (LR)	Stakeholder Consultations
1. Assess the validity of different formative assessment tools and any evidence that their application has led to improved teaching practice and child development outcomes;	RQ 1. Is there any psychometric information about the Formative Assessment (FA) tool? RQ 2. Is there evidence that using the FA tool impacts on child outcomes? RQ 3. Is there evidence that use of the FA tool resulted in a change in teacher practice?	RQ 3. Is there any psychometric information / evidence provided for the FA tools being used? RQ 4. Is there evidence that the use of the FA tool impacts on child outcomes? RQ 6. Is there evidence that use of the FA tool/s resulted in a change in teacher practice?
2. Identify the extent to which any of the effective tools have been currently employed in the NSW ECE sector;		RQ 1. What formative assessment (FA) tools are being used in day-to-day practice, in the NSW ECE sector?
3. Indicate the extent to which the observed outcomes of effective formative assessment tools align with the EYLF	LR Evaluation Framework: assessment of whether the tool addresses each of the EYLF Learning Outcomes (Appendix B and Appendix C)	RQ 4. Is there evidence that the use of the FA tool impacts on child outcomes? (EYLF Learning Outcomes)
4. Indicate whether the tools have been demonstrated to be effective for students with additional needs	LR Evaluation Framework: assessment of whether the tool has been developed and evaluated for use with students with additional needs (Appendix C)	RQ 5. Have the FA tools been demonstrated to be effective for students with additional needs?
5. Consider options for effective tools that can be embedded in the NSW ECE sector, including by providing information on best-practice models of delivery of professional learning and effective tools for the sector.	RQ 4. Is there any evidence that Professional Development supports fidelity of use of FA tools? RQ 5. What are the supports and barriers to using FA tools?	RQ 2. What are the practice models for integrating effective formative assessment tools? RQ 7. What are the models of professional development they have accessed? RQ 8. What are the supports and barriers to using FA tools?

For the Literature Review, we drew on traditional methods of literature searching and evaluation of formative assessment practices. We extended this search method to provide a broader coverage of professional learning and development in the use of formative assessment. For the Stakeholder Consultations, we drew on principles of implementation science to consult with end-users of assessment tools.

The Stakeholder Consultations were conducted in September - October 2019, which included a 2-week school holiday period.

LITERATURE REVIEW

METHODOLOGY

The literature review involved a structured process of formulating the research questions, compiling relevant search terms, selecting databases, conducting the literature search, and iteratively excluding search results that did not match our inclusion criteria. Research questions were formulated in close communication with CESE and the ECED, and informed the approach taken to our evidence search. The research questions were:

- RQ 1. Is there any psychometric information about the Formative Assessment (FA) tool?
- RQ 2. Is there evidence that using the FA tool impacts on child outcomes?
- RQ 3. Is there evidence that use of the FA tool resulted in a change in teacher practice?
- RQ 4. Is there any evidence that professional development supports fidelity of use of FA tools?
- RQ 5. What are the supports and barriers to using FA tools?

DATABASE SEARCHES

In order to gather all relevant literature on formative assessment in early childhood education we implemented four lines of top-down and one line of bottom-up literature searches. Top-down searches included: (1) formative assessment tools and evidence for their effectiveness in Australia and (2) globally, (3) professional development specific to implementation of formative assessment, and (4) supports and barriers to implementation of formative assessment. The bottom-up search included looking for evidence supporting the use of specific tools that were either known or could be viably used for formative assessment. We referred to published reviews on the use of formative assessment tools commissioned by the Victorian Curriculum and Assessment Authority (Cloney, Jackson, & Mitchell, 2019; Marbina et al., 2015; Verden et al., 2018), and an unpublished report produced for Goodstart Early Learning (Harrison & Wang, 2017) to identify specific tools for inclusion in the database searches.

Four databases were used to conduct both top-down and bottom-up searches: EBSCO (Education Resource Complete, Academic Search Premier, OpenDissertations), ProQuest (ERIC, ProQuest Dissertations & Theses Global, Linguistics and Language Behavior Abstracts (LLBA), Sociological Abstracts, Social Services Abstracts), PsychINFO, and INFORMIT (A+ Education, Humanities, and Social Sciences, Families and Society Collection, and Indigenous Collection). In addition, Google Scholar was used to conduct bottom-up searches as it is more inclusive of grey literature and can find more citing studies. Four searches of abstracts in academic databases were conducted.

Search terms were developed and evaluated by the team experts. We formulated groups of keywords around relevant topics and combined them to form each search. We initially formulated an additional list of keywords relating to the EYLF Learning Outcomes, but they proved to be too limited and so were left out of search queries. The search queries we used to conduct top-down and bottom-up searches within academic databases are provided below. Specific keywords for the searches are provided in Appendix A.

- Australia FA tools and evidence for their use: [Formative Assessment] AND [Assessment Tool] AND [Early Childhood] AND [Australia].
- Global FA tools and evidence for their use: [Formative Assessment] AND [Assessment Tool] AND [Early Childhood].
- Professional development on formative assessment: [Formative Assessment] AND [Assessment Tool] AND [Early Childhood] AND [Professional Development].
- Supports and barriers to implementation of formative assessment: [Formative Assessment] AND [Early Childhood] AND [Professional Development] AND [Supports/Barriers].
- Evidence for specific tool use: ([Formative Assessment] OR [Assessment Tool] OR [Professional Development]) AND [Early Childhood] AND [Tool Name]

SEARCH RESULTS

Database searches for evidence of FA tool use and relevant PD studies yielded a total of 1686 abstracts. These references were exported to Endnote X9.2 for further analysis. After removing duplicates and studies clearly irrelevant to our research questions (based on title), 798 abstracts remained. These underwent the full screening process described below. 77 studies passed the screening process and were reviewed in full.

Tool specific searches were conducted in the four academic databases and Google Scholar. We identified 44 tools that matched our criteria were searched. We located a total of 5868 abstracts and 141 were kept for in-depth review after the initial scan.

SCREENING: INCLUSION CRITERIA

Seven criteria were used to evaluate the studies for inclusion in the literature review:

1. **Study Design and Quality**
Due to the expected scarcity of literature, we included any study that provided new data relevant to our questions. We excluded theoretical and commentary papers, as well as proposals for tool development.
2. **Language**
Studies needed to be published in English.
3. **Timeframe**
Studies published between January 2009 and October 2019 were included. The search was restricted to studies published in the past decade to ensure that the technologies discussed were relevant to current classrooms and school contexts.
4. **Student Sample Age Range**
Formative assessment studies of preschool children aged between 3.0 and 5.0 years old (year before school) were included. Studies focusing exclusively on special needs children were excluded.
5. **Tool Properties**
Screeners and other diagnostic tools were excluded. The tool had to be either designed for formative assessment or could be validly used at least 3 times a year for monitoring progress.
6. **Administration**
We only evaluated tools that could be administered by teachers within a realistic time frame (i.e. viable to use in an NSW preschool classroom at least 3 times a year).
7. **Subject Area and Educational Context**
Selected studies needed to focus on formative assessment (or assessment for learning) and its impact on teaching practice or student learning outcomes. Separate searches were undertaken to identify studies exploring optimal school and education system structures, supports and conditions for effective implementation of formative assessment practices.

Abstracts were screened for matching the inclusion criteria. If studies could not be unambiguously excluded, they were kept and passed on to be evaluated by other reviewers with more expertise in education. If the study met the inclusion criteria or could not be unambiguously excluded at this stage, a full text of the study was retrieved and reviewed in depth. Studies that met the inclusion criteria at this stage were evaluated according the criteria described in the next section.

EVALUATION FRAMEWORK

After identifying the relevant studies, each of the retained papers was evaluated by a member of the research team against an evaluation framework and relevant information about each FA tool was entered into an Excel spreadsheet. Three versions of the evaluation framework (spreadsheets) were created:

- to describe tool quality,
- to evaluate evidence for the effectiveness of a specific tool, and

- to evaluate professional development studies focusing on a specific tool.

The frameworks included the following information:

- Study design and quality
Country where the study was conducted.
- Study sample
Sample characteristics including child age, inclusion of special populations, education setting (e.g. preschool, day care, Head Start), study type.
- How the tool was administered
Was special training necessary, who administered the tool, how long does it take and how often is the tool administered?
- Feedback
What feedback does the tool provide, how do the teachers act on the feedback?
- Content
Whether the tool addresses each of the EYLF Learning Outcomes.

For each tool reviewed, we mapped the domains or specific items covered by the tool against the skills described in each of the EYLF Learning Outcomes. For example, the Desired Results Developmental Profile (DRDP) was mapped against EYLF Outcome 1 in the following way:

Outcome 1: Children have a strong sense of identity. Children feel safe, secure, and supported	
Example skills from EYLF	Matching items from DRDP
Children build secure attachments with one and then more familiar educators	SED 3: Relationships and Social Interactions with Familiar Adults. Child develops close relationships with one or more familiar adults (including family members) and interacts in an increasingly competent and cooperative manner with familiar adults
Children use effective routines to help make predicted transitions smoothly	
Children sense and respond to a feeling of belonging	
Children communicate their needs for comfort and assistance	SED 2: Social and Emotional Understanding. Child shows developing understanding of people's behaviours, feelings, thoughts, and individual characteristics
Children establish and maintain respectful, trusting relationships with other children and educators	
Children openly express their feelings and ideas in their interactions with others	SED 1: Identity of Self in Relation to Others. Child shows increasing awareness of self as distinct from and also related to others e.g. Compares own preferences or feelings to those of others
Children respond to ideas and suggestions from others	
Children initiate interactions and conversations with trusted educators	SED 3: Works cooperatively with familiar adults, over sustained periods, to plan and carry out activities or to solve problems.
Children confidently explore and engage with social and physical environments through relationships and play	SED 4: Relationships and Social Interactions with Peers. Child becomes increasingly competent and cooperative in interactions with peers and develops friendships with several peers
Children initiate and join in play	SED 5: Symbolic and Sociodramatic Play.
Children explore aspects of identity through role play	

	Child develops the capacity to use objects to represent other objects or ideas and to engage in symbolic play with others
--	---

6. Psychometric information about the tool
Validity, internal reliability, scale structures

Evaluation of the psychometric properties is a crucial stage development of any assessment or measurement tool. Psychometric evaluation lets users know whether the tool is appropriate for the population of interest, whether the tool measures the intended construct, and whether the measurement remains accurate regardless of who administers the measure. Tools with poor psychometric properties cannot be relied upon for accurate measurement and will likely be unhelpful in facilitating effective decision-making.

Some of the key psychometric information evaluated for this review was:

- Internal reliability – do all the items of the tool equally contribute to the final score? Internal reliability is typically evaluated with Cronbach’s alpha, whereby a coefficient $> .7$ is considered acceptable and a coefficient $> .9$ is considered excellent.
- Inter-rater and test-retest reliability – does the tool produce similar results for a given child when scored by different people or when scored by the same person multiple times? This is typically evaluated through Pearson’s correlations, Kappa coefficients, or joint probability of agreement between scores produced at different time points or by different raters. Coefficients over $> .40$ are considered moderate, and coefficients $> .80$ are considered strong.
- Construct validity – the pattern of correlations with other established measures of the construct. It is expected that the scores will correlate more strongly with other measures targeting the same construct and will have weaker or negative correlations with measures of opposing traits or unrelated measures (discriminant or divergent reliability).
- Structural validity – typically evaluated with exploratory or confirmatory factor analyses, structural validity confirms that the tool has the expected number of dimensions. For example, if the tool has two subscales targeting different constructs, it is expected that it will have two dimensions. Tools that do not conform to the expected dimensional structure need to be revised by either moving items between domains or by changing the scoring to reflect the real number of constructs measured by the tool.
- Developmental and subgroup validity – it is expected that the tool designed for a certain age group will work equally well for children of all ages within this age group, and no items are more difficult for one sex, ethnic group, or other grouping such as second language learner. This is usually measured by differential item functioning. If some items are easier or more difficult for a subset of children with a shared demographic characteristic, the item needs to be excluded or revised. Item Response Theory (IRT) is used to examine variation in difficulty of assessment items and appropriate progression of assessment items within domains and across age groups (i.e., the items are not disordered in terms of expected developmental progression).

7. Details of the professional development

How long and intensive the program was, who are actual and intended users, whether PD addressed specific tool use, provides actionable steps to follow up tool use, and whether classroom materials are available.

8. Outcomes

Both child outcomes and teacher outcomes were recorded where available (but child outcomes were the main focus of the evaluation).

9. Additional comments about the study

Any further evidence from published studies

SUMMARY OF FINDINGS

A total of 22 tools used for formative assessment, with varying degrees of research or technical evidence, were described in one or more of the three Excel spreadsheets. With the context described above in mind, in the next section we present details for four of these tools, each of which were designed to capture growth across a number of domains of learning, which we broadly define as “Domain General Tools”. After describing these domain general tools, we go on to provide details of “Domain Specific Tools” which are designed to assess learning in one domain. We provide detailed examples of two tools (details of all tools can be found in Appendix C).

For each of the selected tools we provide a statement of purpose, professional development and training for use of the tool, potential associations with the EYLF Learning Outcomes, and information organised under each of the research questions, where available:

- RQ 1. Is there any psychometric information about the Formative Assessment (FA) tool?
- RQ 2. Is there evidence that using the FA tool impacts on child outcomes?
- RQ 3. Is there evidence that use of the FA tool resulted in a change in teacher practice?
- RQ 4. Is there any evidence that professional development supports fidelity of use of FA tools?
- RQ 5. What are the supports and barriers to using FA tools?

However, with regard to RQ2 (Is there evidence that using the FA tool impacts on child outcomes?) we were not able to find any studies from which to make a definitive statement of impact. Many studies using the identified formative assessment tools are part of a broader evaluation of curriculum implementation where formative assessment/progress monitoring are just a small part of the package used for individualised programming. This makes it impossible to evaluate the direct contribution of using a formative assessment tool as distinct from many other elements of the curriculum implementation including extensive professional development to support curriculum implementation and the engaging of wider support structures in the community (e.g., family engagement and support advocates). There are clear examples of this provided in the “leading by exemplar” case studies reviewed by LeBetti (2019). We therefore do not address RQ2 further in the examples below.

DOMAIN GENERAL TOOLS

The selected Domain General Tools are aligned with the principles outlined in the Early Years Learning Framework for Australia (DEEWR, 2009) and the Educators’ Guide to the EYLF (DEEWR, 2010) that assessment be ethical, dynamic, forward thinking, child oriented, authentic, and collaborative. All of the measures rely on observation of child development and learning in authentic activities and some form of written recording (ranging from jottings to complete learning stories).

1. TEACHING STRATEGIES GOLD

(HEROMAN, BURTS, BERKE & BICKART, 2010) <https://teachingstrategies.com/solutions/assess/gold/>
<https://teachingstrategies.com/our-approach/research/>

Purpose / Description

Teaching Strategies GOLD (TS GOLD) is associated with the Teaching Strategies Creative Curriculum. It is currently the most widely-used performance-based, observational measure in ECE programs in the US, being required (or actively promoted) in nine state preschool programs (Schilder & Carolan, 2014), and other states being at various phases of implementation of the measure (Weisenfeld, 2017). Results from a survey of 73 Head Start and Early Head Start program directors revealed that 60% of programs were using some form of the TS GOLD assessment system (Isaacs et al., 2015).

TS GOLD is an observation-based teacher rating evaluation instrument designed to help teachers create a developmental profile for each child in order to scaffold his or her learning. TS GOLD makes use of an online platform where teachers score children’s development along 38 different objectives for development and learning. Measured items in the major developmental and learning areas are based on empirical studies of

child processes that are predictive of later success. The progressions in the measure allow teachers to document small increments of progress along a developmental path toward expected milestones. Assessment techniques provide opportunities for teachers to focus on process and to document progress in several areas simultaneously as they observe and interact with children engaging in daily activities. Collected documentation evidence (e.g., observations, artefacts, video recordings, portfolios) is summarized at three checkpoints throughout the year. The measure is intended to be inclusive of ELLs and children with disabilities as well as typically developing children and those who demonstrate competencies beyond developmental expectations.

The purpose of the instrument is to assist teachers in planning appropriate experiences, individualizing instruction, and monitoring and communicating child progress to families and other stakeholders. However, guidelines are vague regarding how educators use the data to inform instruction, and simply state that teachers should, “use the assessment information to tailor instruction to the individual needs of each child” and that the assessment data should be discussed with the teacher assistant to plan for each child. There is feedback on suggested activities and changes in instruction if the assessment tool is used in alignment with Creative Curriculum.

Professional Development

Before using TS GOLD, teachers must complete implementation training (2 days) provided by Teaching Strategies. Training focused on an overview of the measure and an examination of the objectives and child progressions for development and learning (birth through kindergarten). Teachers watched video clips, examined artefacts, evaluated child portfolios, and participated in large-group discussions related to assessment items. They also completed family conference forms and practiced uploading documentation samples, observational notes, and entering progress checkpoint data online. Through an online process, educators can obtain interrater reliability certification. Rigorous initial training and as well as booster training is recommended. Training needs to emphasize the individual areas of development and learning as assessed by particular items, how different domains influence and are influenced by other areas, and how they are intertwined (Becker et al., 2014) as children go about their daily activities.

EYLF Learning Outcomes

While linked to the Creative Curriculum, the instrument can be utilized in other ECE programs. Content matches three of the five EYLF Learning Outcomes.

✗ Outcome 1: Children have a strong sense of identity	✗ Outcome 2: Children are connected with and contribute to their world	✓ Outcome 3: Children have a strong sense of wellbeing	✓ Outcome 4: Children are confident and involved learners.	✓ Outcome 5: Children are effective communicators
		Building relationships with others Interacting appropriately in social situations Gross-motor development Fine motor strength and co-ordination	Approaches to learning (attention, curiosity, initiative, flexibility, problem solving) Memory Classification skills Use of symbols to represent objects, events or persons not present Number concepts and operations Spatial relationships and shapes	Understanding and using language to communication or express thoughts and needs Phonological awareness Alphabet, print and book knowledge Comprehension Emergent writing skills

			Measurement and comparison Pattern knowledge	
--	--	--	--	--

RQ 1. Is there psychometric evidence about the FA tool?

Internal reliability	Inter rater reliability
<p>✓ All of the item, person, and Cronbach's alpha reliability coefficients across all time points and scale scores were above .90 (Lambert et al, 2015).</p>	<p>✓ Inter-rater agreement between master rater and teacher was high. All were above .80, and all but one were above .90 (Lambert et al, 2015).</p> <p>✗ Individual versus class level variance: 38 to 54% of the variance between children was attributable to what classroom the children resided in at the beginning of the year, suggesting teachers' scores of children's readiness skills may reflect information (including teacher bias) that is independent of children's skill levels (Lambert et al., 2015). Other studies report lower levels of 'error' variance (16-25%, Lambert et al., 2014).</p>

Construct validity			
Convergent validity	Divergent validity	Factor structure	Developmental and subgroup validity
<p>✓ Multiple studies report moderate positive correlations with norm-referenced achievement instruments for language, literacy, numeracy (Soderberg et al., 2013; Lambert et al., 2013), and for social functioning and learning behaviours (Teaching Strategies LLC, 2013). Kim et al., (2013) reported evidence in support of uni-dimensionality of each domain.</p>	<p>✗ Significant associations are found across domains (i.e., TS GOLD literacy was highly associated with a direct assessment of math skills) suggesting questionable discriminant validity (Lambert et al., 2013).</p>	<p>✓ Confirmatory factor analysis supports the existence of a 6-factor structure. Longitudinal analysis showed scalar/strict measurement invariance indicating that TS GOLD measures the intended constructs equivalently across time (Lambert, Kim & Burts, 2015).</p>	<p>✓ Teachers can make valid ratings of developmental progress (correlations > .67 with age; Kim et al., 2013). Instrument is sensitive to age differences and growth over time (Lambert et al., 2014).</p> <p>✓ With the exception of 2 language and literacy items, all items function equally well regardless of subgroup membership (e.g., disabilities, English not first language) in a sample of > 50,000 children (Kim et al., 2013).</p>

RQ 3. Is there evidence that use of the FA Tool resulted in a change in teacher practice?

A small number of qualitative studies have examined educator use of TS GOLD in the classroom. Kim (2016) conducted a study involving in-depth interviews and classroom observations of educators. In practice TS GOLD resulted in a more rigid/standardised view of child development from an early age. Kim provides examples of teachers talking about children in terms of levels of achievement within the TS GOLD system. Age-specific, color-coded bands impose homogeneity in children's development and learning by showing "where they're supposed to be" at a certain age. TS GOLD establishes individual differences as deviation from universal norms of sameness and at the same time aims to eliminate these differences.

Teachers prioritized certain play types or playful experiences that yielded useful information about TS GOLD objectives or helped children achieve those objectives (Collado, 2016). Therefore, teachers did not simply rely

on the naturally occurring, spontaneous child observation opportunities. Instead, teachers intentionally organized play settings and planned some activities specifically to observe behaviour they needed to report in the assessment tool.

Collado (2016) conducted a collective case study which included interviews and observations with 6 teachers. Teachers commented that collecting all the data required by the district became 'mechanical and meaningless'. Filling out TS GOLD for all children was viewed as a monotonous process and teachers thought it no longer reflected child's development adequately. While the practitioners agreed that some skills were easier to capture in a checklist, they valued the descriptive information about student learning captured in anecdotal formats. The team described an ideal assessment system as one that prioritized descriptions of students' learning and development as a fuller picture of the student, valuing them as individuals. The [open-ended] questioning was described as a natural part of their daily interactions with students but not recorded, therefore often pushed aside as they were increasingly required to record skill-based data. This frustrated them as they valued shifts in students' thinking over the course of the year. Practitioners studied by Collado (2016) felt strongly that the district and school requirements for data focused too heavily on products rather than the learning process, and that the requirements for data narrowed their instruction, creating a great deal of pressure for teachers. Teachers also shared that they had trouble finding a good way to summarise and visually display all the collected information.

For three of the four classrooms studied by Collado (2016), the teachers realized over time that it was helpful to create a chart with columns for different skills being addressed in assessment activities or play that checked off skills with space to write anecdotal information about how the student performed the skill. Unstructured formats for collecting information about students varied by teacher but served a similar purpose, to record descriptive information about student learning preferences, processes, novel learning experiences, and needs. Teachers continually devised their own assessment tools and tweaked them, but it did not generally (5/6 teachers) feed into properly reviewing the data and planning instructional change.

Participants in Collado (2016) explained that assessments in and of themselves were not effective in revealing all that a student understood and was capable of doing. Therefore, by engaging the student in repeated prompting, modelling, and questioning, the teacher was able to gain a better sense of the student's true abilities and more confidently judge the student's level of performance in that standard. This raises an important point about a lot of time being spent collecting data and little being left to review and plan ahead. A lot of low-level instructional decisions are naturally made on the fly (e.g. series of increasingly specific prompts to help child answer a question/complete a task) - data-collection systems generally won't capture that (and teachers express frustration having to record things like that). Teachers also pointed out that student behaviour appears to have worsened coinciding with the shift towards more data-collection. Teachers relied on developmental progressions and tacit knowledge about what to expect from a child at a particular age to guide their assessment and instruction.

Little et al (2019) conducted interviews and surveys with school district co-ordinators, state officials and teachers in North Carolina. The state provides a pre-approved list of FA tools, and all 6 counties in this study were using the Creative Curriculum and the aligned GOLD assessment system. State guidelines were vague on how FA should be used to inform instruction. From the interviews with administration, it appears that teachers use FA tools with fidelity (daily/weekly data and documentation entry in the electronic system), but there are mixed reports as to whether they actually engage with the tool. Only one county gathered all of FA data together to make county-level decisions and generally the interviewees did not focus on the decision-making aspect. Counties varied in whether they transferred any of their PreK data to kindergarten and how detailed this information was. Only 53% of teachers reported receiving PD specific to assessment in the last 12 months, and only 47% participated in PD about using data from student assessment. 92% of respondents reported using assessment results to help them individualize instruction, and 80% of respondents reported using assessment results to identify individual learning needs.

RQ 5. What are the supports and barriers to using FA tools?

One of the main challenges to using TS GOLD appears to be the time needed to collect anecdotal notes on many assessment items, particularly when the educator is responsible for a large number of children in the classroom. Kim (2016) reported that teachers often collected multiple anecdotal notes per child each time because their assumption about irreversibly progressive child development led them to keep recording observation until they eventually found the "right one". In a later study (Kim, 2018) teachers noted that they struggled to keep up with the recommendation to enter data daily. While TS GOLD advises to find time throughout the day, in practice teachers had to take time out of their break to enter the data or to take the

work home. As they fell behind, they rushed to complete the assessment near the deadline and had to request substitute teachers so that they can have time off the floor. Furthermore, where the system and data entry are monitored by supervisors, teachers felt increasing pressure to keep up with the workload. Despite these challenges, Kim (2018) reported that teachers viewed TS GOLD as beneficial, as it can serve as proof of the work they were doing and showcase the progress their children were making.

2. CHILD OBSERVATION RECORD (COR).

Initially developed in 1993 by Highscope Educational Research Foundation.
<https://highscope.org/cor-advantage/>

Purpose/Description

Teachers or caregivers spend a few minutes each day writing brief notes or “anecdotes” that objectively describe significant episodes of young children’s activities. The anecdotes are then classified and scored according to various COR categories, items, and levels, providing a comprehensive portrait of each child’s developmental gains and the progress of the group as a whole. The measure is designed to capture developmental trajectories of all children.

The COR is administered mainly as a web-based online assessment (including a tablet version). Teachers jot down anecdotes or short observational notes on small notebooks, sticky notes, or directly into the COR application. Along with these anecdotes, teachers can digitally document a child’s words/actions (e.g., audio or video recordings), writing and art samples, and other examples of his or her work (e.g., through digital photographs), and upload them onto the COR Advantage site as part of each child’s portfolio. The observational time period usually lasts 2 to 3 months and is determined by program administrators. A variety of automatically generated reports helps teachers intentionally plan their instructions. This includes growth profiles for each child and reports on group progress. The group reports can be aggregated at various levels (e.g., state, region, program, site, and classroom). To promote family engagement, families can have access to a “parent account” from which they can contribute anecdotes and view their child’s uploaded photos and work. There is a library of developmentally appropriate activities aligned with each COR Advantage level. These resources are available for users as ideas to support children’s development.

Professional Development

To ensure that teachers are reliable observers and scorers, COR training includes how to navigate the online site, what items/ domains and levels mean, how to write and score objective anecdotes, and hints and tips for collecting anecdotes throughout the day. Other pertinent topics covered during the training include how to generate and use various reports (e.g., developmental summary for each child and classroom), importance of complete data/scores to ensure reliable results, and how to share information with families through the family report, Your Child’s Developmental Profile.

EYLF Learning Outcomes

While linked to the High Scope Curriculum, the instrument can be utilized in other ECE programs. Content matches all five EYLF Learning Outcomes.

✓ Outcome 1: Children have a strong sense of identity	✓ Outcome 2: Children are connected with and contribute to their world	✓ Outcome 3: Children have a strong sense of wellbeing.	✓ Outcome 4: Children are confident and involved learners.	✓ Outcome 5: Children are effective communicators.
Social and emotional development: Emotions Building relationships with adults Building relationships with other children Community	Science and Technology: Observing and classifying Experimenting, predicting and drawing conclusions Natural and physical world Tools and technology Social studies: Knowledge of self and others Geography History	Social and emotional development: Emotions Building relationships with adults Building relationships with other children Community Conflict resolution Physical Development and Health: Fine and gross motor skill Personal care and healthy behaviour	Approaches to learning: Initiative Problem-solving Reflection Mathematics: Number and counting Shapes and spatial awareness Measurement Patterns Data analysis Creative arts: Art Music Movement Pretend play	Language, literacy, and communication: Speaking, listening and communication Phonological awareness Alphabet knowledge Reading Book enjoyment and knowledge Writing

RQ 1. Is there psychometric evidence about the FA tool?

Internal reliability		Inter rater reliability	
✓ Content experts supported the usability of COR for teachers and agree that it accurately assesses key development domains (Wakabayashi et al 2019; Waterman et al 2012).		✓ After appropriate training, teachers attained high levels of agreement with experts across all children and items (Wakabayashi et al 2019; Waterman et al 2012).	
Construct validity			
Convergent validity	Divergent validity	Factor structure	Developmental and subgroup validity
✓ Children’s scores on COR were found to correlate highly with the relevant subscales of Woodcock-Johnson III Tests of Achievement and Social Skills Improvement System.	✗ There are disagreements regarding the factor structure due to high correlation between domains. Barghaus & Fantuzzo (2014) present a 4 factor model of social engagement, cognitive skills, coordinated movement and scientific process skills. Possibility that a second order factor common to all items.	✓ Multidimensional Rasch models to evaluate the structural and substantive aspect of construct validity support the theorized eight dimensions underlying the COR Advantage and suggests that scoring rubrics function as the instrument developers intended (i.e., as an eight-level rating scale).	✗ No investigation of differential item functioning across various contexts. ✗ Barghaus & Fantuzzo (2014) using IRT (Item response theory) revealed some items have skill points that do not indicate a developmental progression. ✗ For Head Start teachers applying the COR average assessor variance is 27.6%, indicating that only 70–80% of score

			variation is child centered (Waterman et al., 2012).
--	--	--	--

We have found no evidence to address RQ's 3 to 5.

3. DESIRED RESULTS DEVELOPMENTAL PROFILE (DRDP)

<https://www.desiredresults.us/drdp-forms>

Purpose / Description

This formative assessment instrument was developed by the California Department of Education for young children and their families, and used to inform instruction and program development. The overarching goal guiding the development of the DRDP (2015) was to have one overall assessment approach for all children in ECE programs that provided flexibility in how the assessment is used to ensure developmentally appropriate assessment. It is implemented state-wide in California and Missouri (approximately 500,000 children per year).

Ten key principles guided development of the tool. The tool should:

1. provide a general orientation to facilitating development and learning in key domains at each stage and age;
2. focus on qualitative differences within development in major developmental domains;
3. focus on the child's current level of development rather than on what the child has not yet mastered, while at the same time taking interest in "emerging" or partially mastered knowledge and skills that can further contribute to the curriculum planning process;
4. help teachers track the developmental progress of individual children;
5. be universally designed for use with all children;
6. be reflective of children's cultural and linguistic experiences;
7. be completed by teachers and adults who are familiar with the child;
8. consist of observing naturally occurring learning and behaviour rather than setting up situations to observe and record a child performing an isolated skill;
9. promote an observation and documentation process that supports, rather than interferes with, classroom daily interactions, routines, and activities;
10. consist of a continuum that spans early infancy up to kindergarten entry.

The tool was developed with principles of Universal Design in order to enable access to the assessment for diverse populations. It is specifically designed to allow dual language learners to demonstrate their knowledge and skills. During instrument development, items were excluded if there were potential construct irrelevant cognitive, sensory, emotional, and physical barriers, the items were reviewed for potential sources of bias related to children with disabilities, and the tool was designed to be amenable to accommodations. The tool is appropriate for use with children from the broad range of cultural and language backgrounds. It is supported by additional resources for linguistically diverse assessors.

The DRDP (2015) instrument operationalises developmental continua from early infancy to kindergarten entry into eight constructs that represent domains of development. Each domain is assessed using multiple measures, and each measure consists of a sequence of developmental levels or a progression along which a child's observed knowledge, skills, and behaviours are assessed.

The process for using the DRDP (2015) begins with reflection about children's current knowledge and skills based on evidence collected in natural settings (e.g., ECE settings or home). Evidence of children's knowledge and skills is gathered through teacher observations of children's behaviours, family observations communicated to teachers, and examples of children's work. The measures of the DRDP (2015) are presented in a simple and straightforward manner, clearly demonstrating how learning and development in each area typically progress from early infancy to kindergarten entry. Each measure consists of multiple parts, including a definition, developmental levels, descriptors, and examples, to guide the users in their selections of a rating option for that measure. Definitions specify the aspects of development described in each measure's continuum. The levels within each measure represent ordered and qualitatively distinct points along the developmental progression or continuum that range from earlier to later levels of development.

Descriptors define the behaviours, knowledge, and skills characteristic of typical child development at each level. Examples focus the raters' attention on specific observable behaviours that a child may demonstrate to indicate mastery of a particular developmental level.

Each item is rated on a continuum of skill development (not yet exploring, exploring, developing, building, and integrating). The number of items assessed varies depending on the comprehensiveness of the required assessment (comprehensive, fundamental, essential and snapshot; maximum = 52 for the comprehensive preschool version).

Reports about children's developmental progress are designed to facilitate curriculum planning and inform program and support services for individual children and for groups of children, including the development of goals that are included in IEP planning. Reports are produced that support communication with family members about individual children's learning and development. An interactive website provides aggregated summaries of DRDP (2015) assessment results in relation to child outcomes.

When used formatively, the DRDP (2015) facilitates the production of developmental profiles for each child and for groups of children across the major domains of learning and development. It is designed for teachers to observe, document, and reflect on the learning, development, and progress of all children in an early childhood setting. Within these contexts, the DRDP provides teachers and special educators with information about what children know and can do in early childhood settings and within adult-planned learning activities. Reports are designed to support the use of assessment results for (a) curriculum and program planning and (b) development of IFSP outcomes and IEP goals. The Reports Development Group also developed written guidance to support interpretation of the reports by teachers and administrators.

When used summatively, the DRDP facilitates aggregate reporting of how well children's cumulative learning and development at a designated point in time.

Professional Development

Krause (2016) reported that educators (study of 10 Head Start and 10 State Preschool teachers) were required by the school district to undertake training for the use of DRDP. In contrast, Moiduddin et al (2014) reported that few teachers had ever received formal training on the DRDP, and those who had received training had attended only one training. Teachers mentioned that they had received some informal support for completing the DRDP, including another teacher or administrator sharing training materials or showing them how to document and evaluate whether a child is exploring, developing, building, integrating, or emerging for each measure. Several said that they would appreciate a formal training on the DRDP to make sure they are using the assessment tools correctly and to their full benefit. The general sentiment among all respondents was that they are eager to learn more, and wish they had more time for professional development activities. Supervisors provided support to teachers on DRDP interpretation, goal setting and lesson planning at monthly or one-on-one meetings.

The DRDP website does offer training, webinars, and free resources for educators and administrators (<https://www.desiredresults.us/>)

EYLF Learning Outcomes

DRDP content matches all five EYLF Learning Outcomes.

✓ Outcome 1: Children have a strong sense of identity	✓ Outcome 2: Children are connected with and contribute to their world	✓ Outcome 3: Children have a strong sense of wellbeing.	✓ Outcome 4: Children are confident and involved learners.	✓ Outcome 5: Children are effective communicators.
Social and Emotional domains: Identity of self in relation to others Relationships and social	History and social science domain: Sense of time Sense of place Ecology Conflict negotiation	Social and Emotional domains: Identity of self in relation to others Social and emotional understanding	Approaches to Learning – Self-regulation: Attention maintenance, engagement and persistence	Language and literacy development: Understanding of receptive language Responsiveness to language

<p>interactions with familiar adults Relationships and interactions with peers</p>	<p>Responsible conduct Visual and Performance Arts: Visual art Music Drama Dance</p>	<p>Relationships and social interactions with familiar adults Relationships and interactions with peers Symbolic and sociodramatic play. Physical Development – Health: Perceptual-motor skills and movement concepts Gross locomotor movement skills Gross motor manipulative skills Fine motor manipulative skills Active physical play, nutrition, safety, and personal care routines (hygiene, feeding, dressing)</p>	<p>Curiosity and initiative Self-comforting Self-control of feelings and behaviour Imitation Shared use of space and materials Cognition and number: Spatial relationships Classification Number sense of quantity Measurement Patterning Shapes</p>	<p>Communication and use of expressive language Reciprocal communication and conversation Interest in literacy Comprehension of age appropriate text Concepts about print Phonological awareness Letter and word knowledge Emergent writing <i>Note: All of these can be in the child's first language. There are 4 additional assessments of English Language and Literacy for children who do not have English as their first language.</i></p>
--	---	--	---	--

RQ 1. Is there psychometric evidence about the FA tool?

Internal reliability	Inter rater reliability
<ul style="list-style-type: none"> ✔ Reliability indices ranged from 0.73 to 0.99, indicating that DRDP (2015) domains and sub-domains all had adequate score reliability. ✔ The separation reliability indices of 0.99 indicated that the developmental levels within each DRDP (2015) grouping of domains were highly distinct. 	<ul style="list-style-type: none"> ✔ 31 unique assessor pairs completed the DRDP assessment for 1 to 4 children per pair. Interrater agreement percentages were calculated for both exact agreement (results ranged from 48% to 81%) and agreement within one rating level ranged from 83 to 98%. For preschool-aged children interrater agreement ranged from 50 to 75%, and agreement within one rating level from 84 to 97%. Exact agreement for domain- scaled ratings ranged from 92 to 97% for preschool-aged children.

Construct validity			
Convergent validity	Divergent validity	Factor structure	Developmental and subgroup validity
<p>? alignment with the child completed adaptive assessment (language and math concepts) correlations were low to moderate (Moiduddin et al, 2014).</p> <p>? Assessments were collected for 126 preschool-aged children. DRDP results were correlated with Expressive One Word Picture Vocabulary Test, Receptive One Word Picture Vocabulary Test, Woodcock-Johnson III Achievement tests, and Preschool and Kindergarten Behaviour Scale (DRDP Technical Report, 2015).</p>	<p>✗ No evidence found.</p>	<p>✓ Factor analysis supported a 5-factor structure of the tool at all three time points (self-awareness and identity, mathematics, social skills, language and literacy, general cognitive; Nguyen et al. 2019).</p>	<p>✓ The peak of the ICC for each level was expected to be above a probability rating of 0.5. The research team determined that all of the DRDP measures—and rating categories—demonstrated adequate functioning. The ICCs were ordered, and the category peaks showed distinct separation across the distribution of ability.</p> <p>✓ More than 92% of children progressed on each of the child outcomes, demonstrating sufficient sensitivity</p> <p>✓ 39-60% of children advanced by at least one developmental level between two measurement points, and no measure was overly easy (>90% children advancing one level between two time points) or overly hard (<10% advancing). This was also true for children with mild and severe limitations.</p> <p>✓ Some support that the DRDP is sensitive to children's progress over the year. Most children moved forward in DRDP rating, but 5 to 9% children received a lower rating in the spring relative to the fall (Moiduddin et al, 2014).</p> <p>✗ Items assessing concepts about print, emergent writing, and phonological awareness did not function equivalently for dual language and non-dual language children (Nguyen et al., 2019).</p>

RQ 3. Is there evidence that use of the FA Tool/s resulted in a change in teacher practice?

Three small qualitative studies reported on teachers' use of the DRDP in a variety of classrooms. Moiduddin et al. (2014) interviewed 7 preschool teachers and 2 supervisors, who reported that they used multiple methods of gathering and assessing information to assign ratings. Most respondents said that they combine direct assessment of the skills addressed by each measure with observing and documenting children's skills in an ongoing way, and compiling and reviewing a portfolio of student work. Some explicitly assessed

children's skills by posing specific questions in one-on-one situations or setting up tasks to assess a specific set of skills. All respondents said that they review a file of children's work as well as documentation from their observations and assessments to help them complete the DRDP. All teachers said they complete the DRDP after school hours and/or at home, spending between 20 to 40 minutes per child.

When asked how they review, interpret, and plan based on assessment data, teachers reported that they drew on a range of resources that fall into one of two categories: (1) sources for instructional content and practices and (2) data on children to inform instruction. A few referenced learning standards; others mentioned personal books and published curriculum packages, or used the internet to find activity ideas. Some indicated that children's own interests drive the selection of topics.

Teachers interviewed by Moiduddin et al (2014) were also asked about the steps in the planning process, from processing results to implementing plans based on those results. Most teachers said they used results to select objectives and plan whole-class activities; some use results to form small groups; only a few respondents said that they use results for individualization. DRDP results can be processed and aggregated as a 2-page individual rating record that summarizes the child's progress. It was found that not all teachers complete the rating record.

Nearly all teachers commented that they do collaborate with colleagues and plan together, especially during monthly meetings. However, respondents mentioned that they have limited time to meet and collaborate with each other outside of monthly meetings.

Piper et al (2013) conducted classroom observations and semi-structured interviews with 9 teachers using the DRDP. In line with the Reggio Emilia approach, teachers were seeking to use authentic observation, rather than set up activities just to observe some aspect of development. Observations were typically written down soon after they had taken place, however filling out observation sheets was often done at home/after hours as there was not enough time during the day.

RQ 5. What are the supports and barriers to using FA tools?

Most respondents in the study of Moiduddin et al (2014) commented that the DRDP is useful for lesson planning, in particular for being able to home in on specific measures and skill areas that the whole class needs to develop, using results from the DRDP. A few respondents stated that a strength of the DRDP is its usefulness for gauging students' progress and organising levelled groups according to children's respective strengths and needs. One respondent found the sample lessons provided on the DRDP website for each measure to be valuable. With regard to challenges of using the DRDP, nearly all of the respondents mentioned that the amount of time required to complete the DRDP presents a challenge. A few said that the results are not useful because they lack nuance, represent a one-dimensional snapshot of a child's progress at a single point in time, or are teacher-generated and therefore highly subjective.

Respondents were also asked to share their thoughts on assessment use in general for informing teacher practices (Moiduddin et al., 2014). Nearly all respondents had positive remarks about the usefulness of assessment tools. They expressed relief and satisfaction in using assessment tools to confirm that children are making adequate improvement. Several said that they liked being able to gauge children's progress and use assessment results to narrow the focus of lessons. Another remarked that it is essential to use standardised assessment tools consistently, in order to get an accurate picture of progress and quality across the program - *"We have to have a tool everyone's using to get the same information. We can't have a quality program unless we have tools to assess the teachers, students, program. We need a universal resource so everyone's on the same page."*

Krause (2016) conducted a qualitative study with 10 Head Start and 10 State Preschool teachers about the support provided to EC classroom teachers to aid in the implementation of DRDP. Many noted that reflecting on DRDP results is challenging due to time constraints, and time off the floor with children aids in reflecting on DRDP results. 60% of participants reported that time is a valuable resource to aid in DRDP implementation.

Teachers reported that the DRDP website provided helpful information in implementing DRDP. 20% expressed a desire for there to be fewer items on the measure and suggested that some of the items may be redundant as they are asking for the same information.

4. LEARNING STORIES

(Carr, 1998a)

Purpose / Description

Learning stories are a widely used technique to assess children's learning in ECE Services, particularly in New Zealand. The learning story approach was originally developed in a series of research studies by Margaret Carr. Teachers observe and write narrative stories, and the teacher reflects on the story using learning dispositions to analyse the learning that has taken place; the emphasis here is on the recognising of learning *dispositions* as the learning outcomes. Often photographs are used for illustration and to make the story accessible to the child, which enhances its potential for revisiting. Learning stories link with the New Zealand ECE curriculum *Te Whaariki*. Five dispositions form the basis of assessment in learning stories, and are aligned with curriculum strands and observed behaviours (see table below adapted from Carr, 1998b). Given the holistic and integrated nature of the curriculum, some argue that it is possible to argue that any one of the dispositions could link with any one of the curriculum strands (Blaklock, 2008).

Curriculum Strand	Disposition	Behaviour observed
Belonging	Courage and curiosity	Taking an interest
Well-being	Trust and playfulness	Being involved
Exploration	Perseverance	Persisting with difficulty, challenge and uncertainty
Communication	Confidence	Expressing a point of view or feeling
Contribution	Responsibility	Taking responsibility

Learning stories are said to be suitable for children of all ages and can be made any time during the day when the child is involved in any type of experience, either as an individual or as part of a group. Learning stories may also record the teacher's involvement in the experience. Along with a summary of what learning has occurred, future action that will build on this learning is also recorded. Publications on learning stories do not provide guidelines on where, when and how often to make learning stories for individual children. Common practice in New Zealand ECE services is to carry out one learning story per month for each child, and the observations on which these learning stories are based typically range from 3-10 minutes (Blaklock, 2008).

Professional Development

The professional development program that was designed to assist teachers to carry out learning stories (Blaklock, 2008; Carr, 1998b) provides little information about how the five dispositions are defined.

Alignment to EYLF outcomes.

Due to their subjective nature, learning stories could potentially address all of the EYLF outcomes, although alignment is difficult to map given that there are no objective skills or outcomes stated. There is some concern that a focus on dispositions may lead to neglect of the development of knowledge and skills.

RQ 1. Is there psychometric evidence about the FA tool?

Carr (1998b) suggested that progress in children's learning is made visible in three ways – stories become longer, stories become wider, and stories become more complex. The interpretive nature of learning stories makes psychometric evaluation virtually impossible for this tool. Learning stories embrace the subjectivity of the observer, but the subjective interpretation of learning at the time of the activity means that objectivity is not sought when first describing and documenting a child's learning experience. Hatherly and Sands (2002) argue that "objectivity is gained through making multiple voices visible" (pg 10), although others question whether this is achievable given the time limitations of educators and given the lack of objectivity in the initial observation. Carr (2001) has suggested that the traditional notion of validity is inappropriate for evaluating learning stories, and talks instead of 'local validity', of staff developing a common view of what the learning dispositions and constructs look like in a local context. However, 'local' is not defined and could mean there are differences between teachers in a classroom, between classrooms in an early childhood

setting, or between centres belonging to one provider. Blaiklock (2013) points out that it is also difficult to design a study that could evaluate the changes in children's learning over time as captured by learning stories. If assessments are to be used to show changes in dispositions, there needs to be confidence in the comparability of teacher judgements over time and across different settings (Allal, 2002). There are also concerns that the situational specificity of learning stories may limit their value for assessment and planning in different contexts.

Unlike the previously described formative assessment tools which can be evaluated through conventional assessment of reliability and validity, it has been argued that learning stories are best evaluated by criteria used more widely in qualitative research. However, Blaiklock (2008) raises the issue that it is problematic to apply such techniques for a qualitative assessment such as learning stories which are carried out by educators in the context of everyday work. Blaiklock acknowledges that it is unrealistic to expect early childhood educators to carry out assessment with the same rigor and verification procedures that are required in qualitative research, but that this in itself creates a dilemma – without such rigor there may be little confidence in the value of what is documented in learning stories.

RQ 3. Is there evidence that use of the FA Tool/s resulted in a change in teacher practice?

Learning stories are generally perceived in a positive light by educators (Loggenberg, 2011; Niles, 2016, Goodine, 2013), by lecturers and students (Nyland & Alfayez, 2012), and by program administrators (Goodine, 2013). Loggenberg (2011) found it to be the most common form of assessment in a sample of 25 educators in New Zealand, followed by written observations and annotated examples of a child's work. The benefits of learning stories include the use of contextualised information and inclusivity (Nyland & Alfayez, 2012, Zhang, 2017), being a memorable documentation of the preschool years, and visually appealing to parents (Zhang, 2017).

Despite these overall positive perceptions, the lack of structure raises concerns about the subjective nature of learning stories (Zhang, 2017), including teachers not covering all domains of learning in their assessment (Loggenberg, 2011) or focusing on one domain over another (Goodine, 2013), and collecting evidence primarily in response to parents' concerns (Niles, 2016). Niles (2016) further reports that the interviewed teachers struggled to find the 'right' way to do learning stories, were uncertain about which stories to include, were conflicted about always framing deficits or learning needs in a positive light, and differed in how specific they made the child's learning goals. Finally, learning stories are not a comprehensive tool, do not allow educators to effectively identify early obstacles that learners may experience and do not allow educators to compare data between assessments (Loggenberg, 2011).

Some of the issues around using learning stories can be mitigated. Hooker (2017, 2019) found that structured e-portfolios into which teachers could write the learning story resulted in more frequent entries, saved time, produced more succinct writing, and promoted teacher reflection and frequent revisiting of learning with children, parents and other teachers. As a result, children became involved in conversations about learning more often, and were also more supported to continue their learning at home. Teachers found e-portfolios to be easier for planning, primarily due to their ease of access and the ability to add additional information to previous records. E-portfolios were accessed through iPads as the children were working, and so the children could also contribute to the writing of their learning stories.

RQ 5. What are the supports and barriers to using FA tools?

The primary limitation of using learning stories is time (Nyland & Alfayez, 2012; Zhang, 2017; Niles, 2016; Loggenberg, 2011; Buldu, 2010). Lecturers interviewed by Nyland and Alfayez (2012) suggested that learning stories are better suited to smaller groups with experienced educators, and especially highlighted the necessity of reflection on what is written. In the centre studied by Niles (2016), teachers were responsible for keeping learning story portfolios for 12-15 children. However, some teachers recorded observations for any child, while others only kept records of 'their' children. As a result of the time spent writing learning stories, little time is dedicated to reflecting on them (Nyland & Alfayez, 2012; Niles, 2016, Loggenberg, 2011). For example, educators interviewed by Goodine (2013) were allocated one hour per week as planning time, but it was not consistently provided due to floor responsibilities. Niles (2016) comments that the common practice in New Zealand is to dedicate two hours for planning, although no formal regulations were in place at the time.

Lack of motivation was identified as the main challenge to using learning stories by Buldu (2010), but teachers using e-portfolios as a template for writing their learning stories were motivated by more regular

communication with parents and other teachers that the format facilitated. Another challenge identified by Buldu (2010) is the lack of technological resources. Goodine (2013) found that providing educators with individual cameras and keeping the portfolio books readily available made the data collection process more streamlined. The books were also made available to children, thus including them in the assessment processes and promoting the development of self-assessment. Keeping the books ‘on the floor’ allowed to overcome some of the time-related limitations as teachers began using children’s rest period and outdoor play periods to document assessment data, collaborate, and reflect. However, the study also introduced educators to a list of reflection questions, a reflection journal, and two articles on formative assessment, and the centre had also recently implemented a developmental checklist. Overall it is impossible to evaluate the real effect of any of these changes, as only three educators were interviewed.

Some argue that other forms of narrative record may provide more detailed and accurate information about a child. Running records are considered to be a more cautious approach – here the teacher would keep systematic running records of the child’s activities (preferably at least five) before drawing conclusions or making interpretations of the data. This running record could also be discussed with other educators or family members in order to gain other perspectives on what learning may be occurring. Carefully made running records could provide a rich source of information that may provide more transparent data compared to that collected in a single learning story.

DOMAIN SPECIFIC TOOLS

Eleven domain-specific tools were identified (see Appendix B). Often these are used to help identify children who might benefit from supplemental intervention and serve as frequently administered progress monitoring tools to evaluate response to intervention. Many of these tools are currently administered by a researcher but are simple enough to be adaptable to educator administration in the classroom. Whilst these tools are developed to be authentic assessments, they do require purposeful set up of activity, as opposed to being part of the typical observations that the educator may conduct in day to day activities. However, such activities may be needed to supplement ‘spontaneously’ observed behaviours and skills if the educator is required to complete a more in-depth domain general checklist like those advocated earlier in the report.

We provide more detailed information of two of these domain-specific assessment tools – these specific tools are described because they have been subjected to some form of psychometric evaluation and have associated studies regarding professional development.

1. INVITATION TO THE BIRTHDAY PARTY (BP)

https://static-content.springer.com/esm/art%3A10.1007%2Fs11858-016-08184/MediaObjects/11858_2016_818_MOESM1_ESM.pdf

Purpose/Description

The BP was designed to measure key features of mathematical proficiency. First is performance, especially accuracy and speed of response. It also measures cognitive processes underlying performance, including strategies of solution and conceptual understanding. Process information of this sort is needed to illuminate the reasons for a child’s performance, to predict future performance, and to effectively guide instruction. The BP also measures comprehension and use of mathematical language, a key aspect of mathematical proficiency. It is currently being used in the Virginia Kindergarten Readiness Program.

The design of the BP tasks was based upon current cognitive science, developmental and educational research (Baroody 1987), particularly concerning “developmental trajectories” of mathematical concepts, skills, and strategies (Sarama & Clements 2009). The technology developed was also designed to provide for the easy recording of children’s responses and can offer useful reports concerning children’s performance. Further, the technology enables the collection of large bodies of data that can be useful for program evaluation both within localities and on a larger scale. Results uploaded to a centralized, secure server can be used to provide accountability by comparing classrooms with local sites as well as larger areas.

The BP was designed to provide three types of assessment. 1) formative: the BP provides immediate and specific cognitive process information that can be used to guide instruction; 2) evaluative: the BP can provide information useful to evaluate the implementation of a variety of curricula (instead of one specific

instructional program); 3) screening: a short form of the BP can be used to provide a screening instrument for identifying children at risk for mathematical difficulties and those who might require comprehensive assessment and intervention.

BP is a direct child assessment. It has been translated into Spanish, so that it can be used with a diverse population. The game-like activities, drawing on familiar and interesting events like birthday parties, have been shown to work with many ethnic groups (Ginsburg, Choi, Lopez, Netley, & Chao-Yuan 1997) and should be appropriate for children with learning or emotional difficulties.

A report provides item level data for each individual student. The total score is based on the accuracy of responses both within a domain (i.e. number/ operations, shape, pattern, and space) and across domains. Strategy and math language use (directly and indirectly relevant) are also provided within this report.

Professional Development

The test authors have created workshops and a website designed to enable a wide range of early childhood professionals, including assessment specialists, teachers, and education directors, to understand the BP, to enrich their views of children's thinking and learning, to administer it comfortably, and to interpret the results and their implications for instruction.

EYLF Learning Outcomes

The Birthday Party content matches the EYLF Learning Outcome 4: Children are confident and involved learners. It focuses on numeracy.

RQ 1. Is there psychometric evidence about the FA tool?

Internal reliability and test retest reliability	Inter rater reliability
<p>✓ Cronbach's alpha coefficients were satisfactory, ranging from .70 to .94 for all measures across all age groups. Test–retest reliability coefficients were highest for the Number and Operation measure across all age groups, with the largest value of .82 for age 4. Generally, the lowest test–retest reliability was observed for the Pattern and Space measures (Lee, 2016).</p>	<p>✓ For the inter-rater reliability measured by shadow scoring, all coefficients were high (>.90) except the Pattern and Space measures for the age 3 group (.71 for Pattern and .81 for Space) (Lee, 2016).</p>

Construct validity			
Convergent validity / concurrent and predictive validity.	Divergent validity	Factor structure	Developmental and subgroup validity
<p>✓ Positive correlations with a standardised achievement task indicated adequate concurrent and predictive validity.</p>	<p>✗ No evidence located</p>	<p>✓ Confirmatory factor analysis confirmed acceptable construct validity for all age groups on an underlying construct of 'early mathematics proficiency'</p>	<p>✓ DIF (differential item functioning) analysis suggest that items are functioning the same way regardless of gender, SES, or language spoken at home.</p> <p>✓ Findings from Classical Test Theory and IRT item analyses indicate that item difficulty displayed optimal variation and increased both within and across age groups on common items. The progression in</p>

			difficulty lends empirical support to the theoretically developed items.
--	--	--	--

RQ 3. Is there evidence that use of the FA Tool/s resulted in a change in teacher practice?

Ertle et al (2016) conducted a PD study with 11 early childhood preservice students enrolled in a graduate mathematics methods course. The BP workshops were designed to meet three goals: (1) train teachers on how to administer the BP; (2) provide teachers with background knowledge and understanding of the skills targeted by the BP; (3) offer instructional next steps based on the assessment results. PD was provided via face to face sessions and through web resources. The current study was a pilot evaluation of the website resources. This pilot study provides preliminary support for the value of this video-based, assessment-driven approach to mathematics PD. The respondents all identified value to the website with regard to conducting assessments, understanding the purpose of assessments, or what they could learn about mathematics, children's mathematics learning, or how to teach mathematics. Results of the pilot study suggested the website was not entirely successful in helping respondents use the website ideas to shape specific teaching practices.

2. INDIVIDUAL GROWTH AND DEVELOPMENT INDICATORS – EARLY LITERACY (EL-IGDIS)

<https://www.myigdis.com/preschool-assessments/early-literacy-assessments/>

Purpose / Description

EL-IDGIs is a progress-monitoring and screening measure that includes five literacy tasks: picture naming (oral language), sound identification (alphabet knowledge), 'Which one doesn't belong?' (comprehension), rhyming and alliteration (phonological awareness). EL-IDGIs were designed using Item Response Theory and are currently being funded and continuously improved. An overall score is provided.

Most validation studies use typical samples (English as first language, typical development, no special education services, not living in poverty). One early study found significant differences in performance for children with speech and language difficulties, children living in poverty, and dual language learners (Missall et al., 2006). Content of the test items is reflective of objects and experiences present in a typical preschooler's environment, and the diversity of the items sampled during the test is expected to control for any cultural biases experienced by individual children.

Professional Development

Only basic training on using the test materials is needed. Publishers advise that the coordinator oversees EL-IDGIs use and interpretation should be done by a professional in special education, early childhood education, psychology, speech and language, school nursing, or another closely related area. The tool takes approximately 10 minutes to administer, 3 times a year. An online platform (myIDGIs) is available for storing assessment data, producing summary reports, and track performance over time.

EYLF Learning Outcomes

The EL-IDGIs content matches the EYLF Learning Outcome 5: Children are effective communicators. It focuses on oral language, alphabet knowledge, comprehension, phonological awareness.

RQ 1. Is there psychometric evidence about the FA tool?

Reliability

Test retest reliability
✓ Hojniski & Floyd (2013) report test-retest reliability as .93–.97.

Validity

Concurrent validity	Developmental and subgroup validity
<p>✓ Roseth et al. (2012) found medium to large correlations between EL-IDGI domains and standardised measures (PPVT-4, TOPEL, TOPEL-PA, TOPEL-PK, CELF-Pre-2), ranging from .52 to .71.</p>	<p>✗ Roseth et al., (2012) report sensitivity of .71–.77 and specificity of .57–.69, implying that we cannot assume linear growth in EL-IGDI data across the entire developmental period of the scale (30–66m), especially for rhyming and alliteration. Rhyming and alliteration measures may have a floor effect, or the skills themselves require greater maturation than the picture-naming measure.</p> <p>Evidence from Roseth (2012) suggests the tool might not be sensitive or that a one-time assessment might not be appropriate given that there are age-related differences in growth rate.</p>

RQ 3. Is there evidence that use of the FA Tool/s resulted in a change in teacher practice?

Gettinger & Stoiber (2012) used 3 of the EL-IDGIs as part of a larger tiered literacy intervention program (EMERGE). The intervention appears to be effective, but it is impossible to separate the role of FA from other intervention components (teacher PD on using the tool, literacy intervention).

PROFESSIONAL LEARNING AND DEVELOPMENT

As noted in the previous sections on Domain General and Domain Specific assessment tools, professional learning (PL), development and training is an essential component of effective assessment practice in ECE. Teachers' ability to reliably and accurately score standardised assessment tools, such as TS GOLD, COR, DRDP, BP and EL-IGDIs, rely heavily on teachers' ability to observe children's naturally occurring words and actions and make judgments about what children can and cannot do. The examples described for each of these tools indicate a variety of models for training. These include: training to develop observer reliability in the use of the tool (e.g., TS GOLD); training to improve skills in observation, recording, and sharing information about the child with families (e.g., COR); access to training materials (e.g., DRDP); workshops to understand, administer and interpret results of the tool (e.g., BP); and informal support, supervision, or mentoring (e.g., DRDP, EL-IDGIs). However, none had been evaluated in the studies that were identified and reviewed, and where PL was inquired about, educators expressed the need for further support with training.

The evidence from these five tools is consistent with studies of professional learning, development and training. A survey of 159 early education educators and specialists found that the majority of respondents were ambivalent about the adequacy of training to conduct and interpret screening, assessment, and evaluation (Gokiart et al., 2013): 80% of respondents wanted to learn more about assessment and how to link results to programming and instruction; 61% of participants reported that they received either no training or 1 day/in-house training on using the tool.

Susman-Stillman et al. (2014) surveyed a large number of early educators and PL providers, finding that both groups attributed low implementation and fidelity of assessment to a lack of knowledge and skills on authentic assessment (AA). 81% of educators rated themselves as competent at conducting AA, but only ~60% felt competent at assessing children with disabilities, and culturally and linguistically diverse children. Many of the PL providers (87%) believed ECEs struggle with implementing AA, and only 52% agreed that ECEs get the support they need from their director/manager/coach to implement authentic assessment. PL providers also felt limited competence at training to assess children with disabilities (76%), culturally diverse children (71%), and English language learners (63%).

These two studies align with an earlier report by Myford and Wolfe (2003) that results from studies conducted over the last three quarters of a century, across a variety of assessment contexts, indicate that not all teachers are skilled observers, and some have difficulty assigning scores in a systematic and defensible manner. To support this point, early childhood researchers have found moderate-to-large scorer variance in

teacher scored observation-based assessments, that is, variations in children's scores that are attributable to the scorer effects, such as leniency or strictness, rather than to the children themselves (e.g., Lambert, Kim, & Burts, 2015; Waterman, McDermott, Fantuzzo, & Gadsden, 2012).

MODELS OF PROFESSIONAL LEARNING

The literature review located a number of studies that had evaluated different approaches to the delivery of professional learning / development and training for ECE educators. Whilst providing useful background information, none specifically focused on PL in the use of assessment tools. Professional learning research on non-assessment elements included partnering with academics or mentors that received support from leadership within the organisational context. In Australia, Hadley, Waniganayake and Shepherd (2015) examined literature on different models of professional learning. They found professional learning offered by academics to ECE services to be an effective model to enact change. This included a strong focus on practitioner inquiry and the importance of having pedagogical leadership supported by the employing organisation. The significance of organisational support was also evident in Nolan, Morrissey and Dumenden's (2013) two year study in which Australian early childhood teachers were mentored by a more experienced colleague in their centres. Teachers reported this model to be highly supportive for not only newly trained teachers but also for teachers working in more isolated contexts. Another Australian study, *Fostering Effective Early Learning* (FEEL), study offered a similar model with academics supporting educators' professional learning during centre visits to develop educators' knowledge in key learning areas such as self-regulation, language and communication, maths, science and critical thinking. The FEEL study showed significant changes in the educators' practice in relation to the quality of curricula and interactions at the centres receiving the intervention (Siraj et al., 2018). FEEL also underlines the importance of the role of commitment by key stakeholders / providers of ECEC to facilitate and maintain professional learning of early childhood staff by allocating funding to support educators in their professional development. However, while these studies demonstrate the success of these PL programs, we were unable to locate any explicit systematic evaluations of PL in Australia that focus specifically on assessment.

PROFESSIONAL LEARNING FOR ASSESSMENT

In order to supplement the literature review and provide specific information on PL for assessment, a Google search was conducted to identify providers of PL courses for ECE in NSW and other states. The search retrieved over 20 organisations, which comprised a diverse range of providers, including for profit, government and community services.

We then reviewed on-line material for each of these organisations using search terms "assessment" and "documentation" to determine whether these services provided specialised training on formative and summative assessment. Results of these searches identified 31 PL offerings, which are summarised in Appendix D. For each PL course we identify the name of the provider/organisation, the title, time commitment, cost and delivery mode of the relevant PL course, location, and whether it is NESA accredited. Additional comments are also included, where relevant.

In sum, there are a number of PL provider organisations, such as Gowrie NSW Education Hub, Semann and Slattery, and Community Early Learning Australia (CELA) that provide day training sessions in Sydney and surrounding areas. A notable lack of availability of PL offerings was found in regional and remote areas. Services in these areas may need to investigate other options such as webinars or custom-tailored training that can be quite expensive, as presenters charge for travel as well as their time for delivery of the training.

PL is also offered through in-house training provided by large organisations, such as KU Children's Services, SDN Children's Services, Goodstart Early Learning, as well as some smaller for-profit services, such as Active Kids and Explore & Develop. These organisations offer face-to-face training involving annual teachers' conferences and centre workshops with a specialist support worker who supports all the staff.

Collectively, these professional learning offerings advocate that PL can improve educators' knowledge and skills through face-to-face, webinars or online self-paced modules. Courses ranged in cost from providing free resources (e.g., Early Childhood Australia; Children's Services Central, Gowrie NSW Education Hub) to more expensive options involving face-to-face training (e.g., \$250 per person, CELA). There was also customised training with private consultancies although no costings were available on the websites.

The search and review process showed a clear need for specifically PL opportunities on formative and summative assessment in ECE across Australia. This was further evidenced by the titles of the courses, none of which directly used these terms. Through strengthening understandings of what these types of assessment looks like, educators may be better able to examine, adapt and articulate formative and summative assessment within their workplaces.

We extended the search for PL offerings to include professional publications designed to support educators. Search terms included topics such as pedagogical documentation, assessment, observation and critical reflection. Some of the identified publications were quite comprehensive in the information they provided. For example, a free 56-page e-book is available from Gowrie SA (Sisson & Whittington, 2018) funded by the Department for Education and Child Development (SA) which encourages educators/teams to explore a range of provocations as a means of engagement with the resource.

It is important to note that we found that searching with particular terms was problematic. Searches using the term “assessment” led to professional learning around the NQS Assessment and Rating process while searches including the term “understanding children” directed us to workshops on guiding children’s behaviour. It would seem that many terms familiar to the early childhood sector have connections to other areas where they are not necessarily related, making it difficult for educators and service providers to find exactly what they want in the way of targeted professional learning on formative assessment.

STAKEHOLDER CONSULTATIONS

A case study approach was used to gather data in eight selected ECE services for children aged 3- to 5-years that had shown exemplary and innovative formative assessment practices, as identified by the National Quality Standards (NQS) Assessment and Rating (A & R) process. Ethics approval for the study was applied for through the Macquarie University Faculty of Human Sciences Subcommittee of the Human Research Ethics Committee, and the Charles Sturt University Ethics in Human Research Committee. Approvals were provided by both Universities.

METHODOLOGY

RECRUITMENT OF ECE SERVICES

An analysis of the NQS database conducted by CESE for the DoE identified 38 ECE services in regional, metropolitan and rural NSW that met the criteria for inclusion in the study; that is an Exceeding NQS rating on Quality Area 1 and Quality Area 7, or other evidence of exemplary practice in assessment, based on Assessment and Ratings (A&R) reports. CESE made an initial approach to each of these services by email to ascertain their willingness to participate in the study. Services that responded within the timeframe were provided to the research team, along with their confidential A&R reports, to finalise the list of services to be invited to participate.

The research team decided on selecting a diversity of services that considered the following characteristics:

- location (metropolitan, regional, remote);
- type of program (preschool; long day care);
- Approved Provider organisational structure (for-profit company, not-for-profit community-based organisation, not-for-profit government provider);
- size of Approved Provider organisation (small, standalone service; medium 2 to 7 services; large 8 or more services).

Selection also considered feasibility issues, including travel time, and whether there were any additional requirements for approval of research by an Approved Provider.

The selected services included an equal number of preschools and long day care centres, located in metropolitan, regional and remote locations. They provided a good representation across all criteria:

- Preschools (2 community-based, 1 church-based, 1 Department of Education school-based)
- Long Day Care Centres (2 community-based, 1 local government-based, 1 for-profit company)

- Type of Approved Provider: 1 for-profit, 7 not-for-profit
- Size of Approved Provider: 4 small standalone; 2 medium (2-7 services); 2 large (>7 services)
- Location: 3 Metropolitan, 3 Regional, 2 Remote

All eight services were approached by a member of the research team, by email and telephone. All eight agreed to participate.

PARTICIPANTS

Invited participants were Centre Directors, early childhood teachers, educational leaders, educators, and representatives from Approved Provider organisations who held responsibilities for assessment and planning for preschool-aged children in the year before school. The number of participants per service ranged from 2 to 6, with a total of 35 who provided interview data. These are described in terms of their roles/positions:

- 3 Approved Provider (AP) representatives
- 8 Centre Directors (CD)
- 7 Educational Leaders (EL)
- 8 Early Childhood Teachers (ECT)
- 9 Educators (Ed), including specialist support staff / liaison officers (SSLO)

DATA COLLECTION

A member of the research team made a one- or two-day visit to each case study service to conduct on-site interviews with the Centre Director(s) and staff who worked with children in the year before school. To ensure that all relevant staff who wished to participate in the interviews were able to do so, funding for backfill was provided. Representatives of AP organisations were interviewed on-site or by telephone.

The purpose of the interviews was to:

- gather perspectives about the relevance, importance and use of formative assessment for children in the year before school;
- gather information and examples of the existing formative assessment techniques and documents;
- consider how formative assessment(s) match the EYLF Learning Outcomes;
- discuss the use of formative assessments in the writing of a Transition to School Statement;
- gather information and examples of how child assessments are shared and discussed with parents, including any parent-completed assessment;
- gather perspectives on the professional development, support, and resources educators / teachers have received or accessed to help them select and use formative assessment tools with children; and
- gather perspectives on possible barriers to the use of formative assessment tools.

A set of interview questions were prepared, reviewed with CESE and the ECED, and transferred to an on-line SurveyMonkey format prepared by the Charles Sturt University Spatial Data Analysis Network (SPAN). Questions included options for single responses (yes, no), ratings (1 to 5) and more in-depth narrative responses.

Interviews were conducted one-on-one, in a quiet location away from the children. Participants' responses were entered directly into the on-line survey or recorded in hand-written notes and subsequently transcribed into the survey.

DATA MANAGEMENT

SurveyMonkey data files were stored and provided to the research team for analysis by SPAN in two formats: Excel and PDF.

DATA ANALYSIS

Prior to commencing analysis, the research team reviewed and made minor revisions to the key research questions that were the basis for our analysis of the interview data. These questions were:

- RQ 1. What formative assessment (FA) tools are being used in day-to-day practice, in the NSW ECE sector?
- RQ 2. What are the practice models for integrating effective FA tools?
- RQ 3. Is there any psychometric information/evidence provided for the FA tools being used?
- RQ 4. Is there evidence that the use of the FA tool impacts on child outcomes?
- RQ 5. Have the FA tools been demonstrated to be effective for students with additional needs?
- RQ 6. Is there evidence that use of the FA tool/s resulted in a change in teacher practice?
- RQ 7. What are the models of professional development they have accessed?
- RQ 8. What are the supports and barriers to using FA tools?

Quantitative Data Analysis

Interviews with 24 EL/ECT/Ed staff and 3 AP representatives provided quantitative, numeric data, in the form of yes/no responses and ratings on a 1 to 5 scale. These data were analysed using simple descriptive statistics to generate percentages of yes/no responses and average ratings on 1-5 scales (1 = least, 3 = neutral, 5 = most, and 1 = poor, 3 = neutral, 5 = very good).

Average ratings (1-5 scale) were also computed for 12 additional ratings collected on specific assessment tools: ease of use, suitability for sharing with families and with other educators, suitability for children with a range of abilities or children who do not have English as their home language, informing the writing of a Transition to School statement, the time needed to complete the assessment, usefulness for assessing children's learning in relation to the EYLF Learning Outcomes, cost, suitability for involving children in the assessment, and usefulness for providing ongoing feedback, reflection and planning for individual children.

The small number of services (8), the different numbers of participants per service (2 to 6), and the selective nature of the sample limits the validity and generalisability. Findings must be viewed with these limitations in mind.

Qualitative Data Analysis

All participants provided qualitative, text-based data. These interviews were analysed in a series of steps. First, data from the CD and EL/ECT/Ed interviews were transferred from the Excel spreadsheet to a word document and tabularised under the interview questions. Next, two researchers independently coded the data (separately for the EL/ECT/Eds and CDs) and then arranged the codes under the over-arching research questions to which the codes corresponded.

Second, the two researchers conferred on the codes and arrangement. There was strong agreement between the researchers on the codes and arrangement. However, on discussion, some codes were shifted under different research questions. This process resulted in a 'coding rubric'.

Third, one researcher analysed the remaining data using the coding rubric, adding new codes where necessary. A similar process was followed for the AP data.

SUMMARY OF FINDINGS

SERVICE PRACTICES FOR CONDUCTING FORMATIVE ASSESSMENT

EL/ECT/Eds and AP representatives were asked about expectations for who should undertake formative assessment, and when and where they recorded and reflected on their assessments of children's learning. Of the 24 participants, 19 stated that a mix of staff recorded assessments, while other five stated that only the ECT or EL did this work. Of the 23 participants who answered the question about how often assessments should be recorded, 11 said daily, five said weekly, and six said at least monthly, but participants also mentioned that frequency varied by the type of assessment or the child's attendance pattern.

Participants reported wide variation in the time they had allocated for writing their assessment records and planning, from 7 hours (1 day) per week to none, depending on their position / role within the service. Most

common responses were 7 hours per week for ELs, 3 – 4 hours per week for ECTs and 1 - 2 hours per week for Educators. An educator with A Certificate III qualification mentioned she had some time allocation, but this was not 'officially' part of their role. Specialist Support / Liaison Officers were the least likely to receive a set time allocation and reported that they were expected to do this in their own time. All participants said their service / school provided with a space to document and plan away from the children.

THEMATIC ANALYSIS

From the analytical process outlined above, the themes that emerged were organised under the over-arching research questions. In the following sections we present these themes, with exemplary quotes (where available) and quantitative data (where available) for three positional groups:

1. staff who work directly with the children: Educational Leaders (EL), Early Childhood Teachers (ECT), Educators (Ed), including Specialist Support / Liaison Officers (SSLO)
2. Centre Directors (CD)
3. representatives from Approved Provider (AP) organisations

DIVERSE TYPES OF FORMATIVE ASSESSMENT TOOLS USED IN DAY-TO-DAY PRACTICE

Data from 27 participants (EL, ECT, Ed, CD plus AP reps) presented in Table 2 indicated that staff in the case study ECE services were using a wide range of different types of formative assessment tools. Between 26% to 96% of participants reported that they used the 12 types of assessment tools identified in the interview questions. Participants were also asked about their personal views of the usefulness of these assessment tools, from least (1) to most (5).

Written observations with and without photographs, and learning stories, which were the used by over two-thirds of participants (67% to 96%), were rated highly on usefulness for formative assessment (Ms > 4.5 on a 1-5 scale). A high proportion of participants (85%) endorsed the use of digital apps as a platform for organising, storing, sharing, and consolidating children's records, but rated them less highly on usefulness for formative assessment (M = 4.2).

Table 2: Responses to Use (yes/no) and Perceived Usefulness (1 – 5 rating) of Assessment Tools

	Yes		No		Rating	
	n	%	n	%	n	M
Digital app/s	23	85.2	4	14.8	22	4.18
Written observations (own words no photographs)	20	74.1	7	25.9	19	4.53
Written observations (plus photographs)	25	96.2	1	3.9	26	4.69
Photographs only	12	44.4	15	55.6	12	3.33
Learning Story	18	66.7	9	33.3	18	4.56
Floor Books	15	55.6	12	44.4	15	3.87
Developmental checklist	19	70.4	8	29.6	18	4.44
On-line assessment tool	7	25.9	20	74.1	6	4.50
Service designed tool	14	51.9	13	48.2	12	4.75
External assessment tool	7	26.9	19	73.1	7	4.57
Child completed assessment	10	38.5	16	61.5	9	4.44
Family completed assessment tool	14	53.9	12	46.2	13	4.54

Interestingly, high ratings for usefulness as a formative assessment tool were given for tools that were used by fewer participants; 'on-line assessment tool' (used by 26%, rating M = 4.5); 'service designed tool' (used by 52%, rating M = 4.75); 'external assessment tool' (used by 27%, rating M = 4.57); 'family-completed assessment tool' (used by 54%, rating M = 4.54).

In addition to these assessment tools, many participants also described “other tools” that were currently in use in their services. These included reflective journals, instruments that had been introduced for an external research study, and tools that were linked to a particular educational program. Only one of the three APs had strong knowledge of the types of FA tools used in their services. There was no expectation amongst APs that educators complete specific assessments. One AP in particular emphasised the importance of FA being responsive to context. CDs saw the use of multiple methods as important to address different aspects of assessment.

Generally, participants agreed that FA tools: (i) are important for pedagogical planning; (ii) can be used to involve children in documenting their learning, and (iii) are an important way for communicating with families – this latter was seen by CDs as being of particular importance but some tools were seen by CDs as better for communicating with families and others for informing teaching. However, the participants’ views varied in how well these aims are achieved via the different tools utilised.

A number of tools mentioned or provided by participants, and other tools less frequently mentioned such as the ECERS-R, SSTEW, and RAPIE (see below), are not designed for ‘formative assessment’. This perhaps suggests some misunderstanding about what FA is in ECE.

In the following sections we provide details of participants’ responses to each of the assessment tools.

OBSERVATIONS

The most common FA tool for day-to-day practice that participants spoke about as being useful for supporting program planning was observations – both group and individual. The type of observation noted as most useful was almost exclusively limited to anecdotal observations and / or jottings of children’s interactions (i.e. other types of observations such as running records and time-sampling were not mentioned) – either hand-written or recorded on a digital device.

Several educators reported hand recording and later transcribing observations onto digital platforms, whilst others recorded their observations immediately onto digital platforms. They included photographs, samples of children’s work, as well as educator reflection focused on children’s learning, skills, and development, and resources required for the future (i.e., they followed a planning cycle).

Some educators mentioned that they focused their observations on children’s individual goals or the EYLF outcomes (see Figures 1 and 2). Others commented that their observations of children were also informed by

their “*knowledge of the child and what parents say*”. CDs also commented that observations were a useful tool for linking learning to EYLF outcomes.

Figure 1: Example of Observation Record – Group of Children

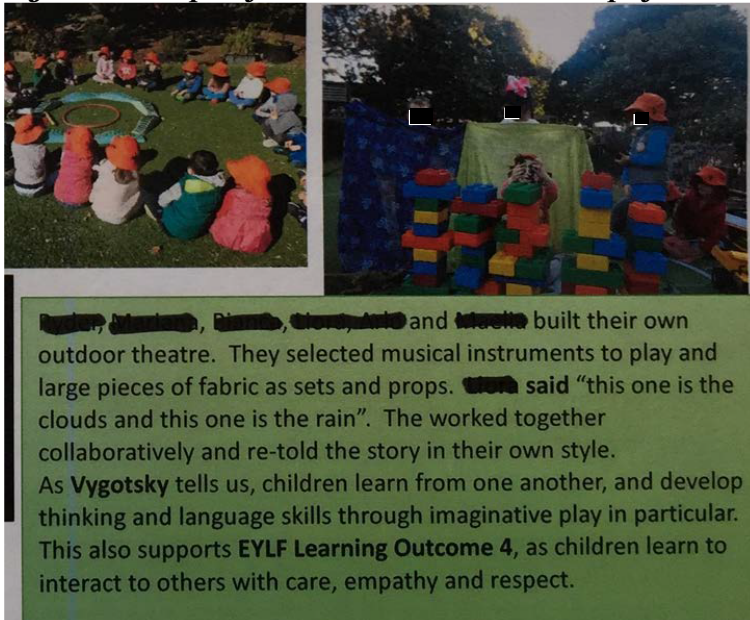
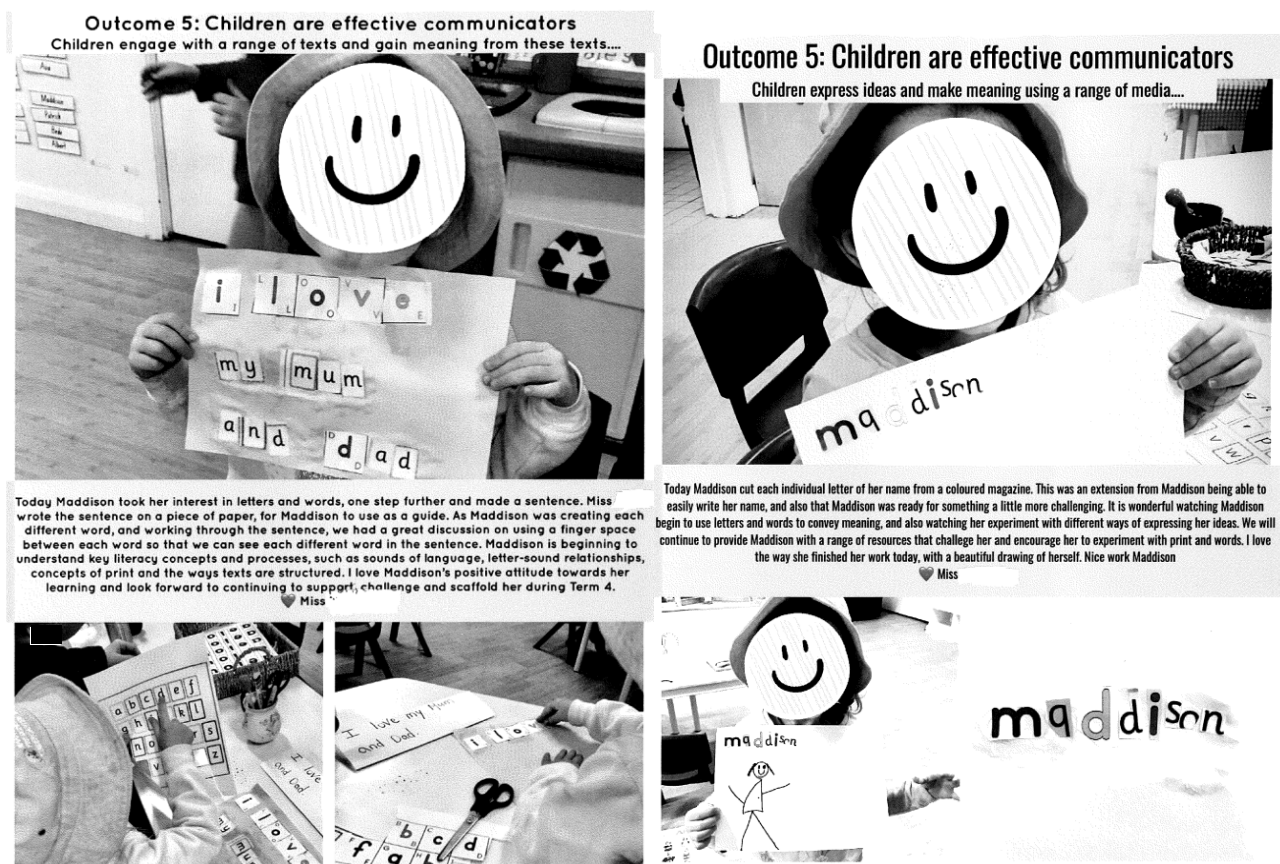


Figure 2: Example of Observation Record – Individual Child



Many educators reported a preference for observations due to (i) their meaningful narrative structure that facilitates sharing of information amongst educators, and (ii) their ability to capture children's activities, routines, interactions, as well as their strengths, accomplishments, dispositions and interests. For example, one educator noted that observations enable the capturing of children's "*capabilities, interest and confidence levels*".

Some educators commented that observations *with* photographs was preferable because photographs give a context for the learning. However, other educators considered that using observations *and* photographs is too time consuming.

Some educators also commented that observations enabled them to look back on children's learning. For instance, one educator commented:

Written observations are good to have to reflect on future learning. You can see the progress the child has made over the year.

For some educators a benefit of observations is that they can be done quickly and easily while working with children, and, for those that use them, typed onto a digital platform. Conversely, some educators found observations time consuming.

Educators, CDs and APs all noted that the effectiveness of observations depends on the detail in the story and this, in turn, depends on the skills of the individual completing them. Indeed, one AP emphasised the importance of quality observations rather than a set number of observations being conducted by educators.

Several educators reported that it was important to them that children's voices and input is included in formative assessment. Observations were considered by some educators as a providing a valuable way to involve children. For instance, in relation to observations, one educator commented on their "*suitability for involving children: children can see photos and talk about what they did*". But educators also noted that this involvement was dependent on having the time to involve children.

Many educators spoke of the need to have tools that they could use to communicate with families about children's learning. Whilst observations were considered by many educators to be mostly for teachers' use, several educators acknowledged that observations provided good information for sharing with families, particularly if they had photos.

The additional ratings participants gave for observations align with these comments. Ratings were provided for observations without photographs by 19 participants and observations with photographs by 26 participants (see Appendix E). Ratings were high (Means > 4) for both methods, but were consistently higher for observations with photographs vs observations without photographs; e.g.,

- for ease of use (M = 4.7 vs M = 4.4);
- suitability for sharing with families (M = 4.7 vs M = 2.8);
- suitability for children who do not have English as their home language (M = 4.7 vs M = 4.2);
- usefulness for providing ongoing feedback, reflection and planning for individual children (M = 4.8 vs M = 4.6).
- suitable for involving children in the assessment process (M = 4.2 vs M = 3.7).

Only time needed to complete the assessment (M = 3.8 vs M = 4.0) and cost (M = 4.3 vs M = 4.5) were rated lower for observations with photographs.

LEARNING STORIES

A sizeable number of participants reported that learning stories was their preferred formative assessment tool. Ratings provided by 18 participants were high (Means > 4.5) for six of the 12 additional criteria:

- suitability for sharing with families and with other educators,
- suitability for children with a range of abilities,
- informing the writing of a Transition to School statement,
- suitability for assessing children's learning in relation to the EYLF Learning Outcomes, and
- usefulness for providing ongoing feedback, reflection and planning for individual children.

However, learning stories received low ratings for the time needed to complete the assessment ($M = 2.8$) and was not highly endorsed for suitability for involving children in the assessment ($M = 3.9$).

Educators noted that learning stories:

- (i) enable the capture of children's learning across all developmental areas and EYLF outcomes;
- (ii) provide consistent information about individual children;
- (iii) show progress in learning;
- (iv) can be used to deconstruct learning; and
- (v) can inform pedagogical planning.

Learning stories were considered by some educators as most useful in the year before school.

Learning stories were attractively presented, bringing together a narrative of play and learning of an individual child or a group (see Figure 3). They were usually accompanied by more than one photo which directly related to the story. They were shared with families sometimes being posted on a wall in the children's room where they could be viewed, sent to families via an app or printed so a hard copy could be placed in the child/ren's portfolio.

Figure 3: Example of Learning Story

THE ZOO 28/8/19



A group of children created their own zoo today, completely child-led with the educators observing how they organised and ran it.

[Redacted]: This is our zoo
 [Redacted]: It's \$2 to get in
 [Redacted]: I'm a cheetah
 [Redacted]: I'm a tiger
 [Redacted]: I'm a pig
 [Redacted]: I'm a cheetah too
 [Redacted]: I'm a tiger

What learning is happening here?

The group started with the two fence panels and dragged the basketball net over to form an enclosure. This was done with the children directing one another, making suggestions and listening to one another. The tunnel made by the basketball net immediately got Roc and Stan making animal noises and crawling around.

There was negotiation around who would be which animal, who was the zoo keeper and who was a visitor. Problems were solved as a group and there was no adult intervention needed to sort out who was doing what and how the zoo should operate.

They decided when to sleep, eat and move around. The zoo keepers sold tickets, took money and corralled visitors around the zoo.

The children demonstrated imagination, collaboration, skills of negotiation and problem-solving, the ability to transfer knowledge to new scenarios and resource their own learning. They developed their language through their knowledge of animals and how zoos worked and learnt from each other as they allowed everyone in the group to contribute equally.

Possible next steps in learning:

* Investigate the habitats, diet and behaviours of the animals the children are interested in.

Similar to what they noted for observations, educators liked the narrative structure of learning stories. They commented that learning stories enable educators to share detailed information about children's learning readily with educators and families. Typical of the comments was this one from an educator:

Learning stories help to share your perspectives as well, so the documentation is more collaborative than the others. This is my favourite style of observation and documentation for children's learning.

Another educator noted that learning stories are:

Useful because it is an interest of the child, their imagination, their abilities. It's nice for families to see what the child has been involved with in the day.

There were, however, a number of challenges identified with learning stories by educators and AP including:

- they are not always done appropriately (e.g. they often do not include ‘next steps’);
- they can be difficult for new educators;
- they can be too long, so parents don’t read them;
- there may be too many words for families from language backgrounds other than English;
- they are not so easy to involve children; and
- they are too time consuming. One educator commented that it was more important to be with children than stressed about getting these done, and so she/he had moved to checklists.

Educators gave some suggestions for the use of learning stories including doing at least one a month per child and creating versions especially for families.

PORTFOLIOS

Although not included in the list of specified assessment tools, or rated by participants, individual child portfolios were a common tool that educators spoke of using. Educators noted that portfolios can include a variety of assessment methods including observations, photos and samples of children’s work. In some cases, particularly when a digital platform was not being used, all individual records were pasted into the child’s portfolio. These were seen as useful for both ‘keeping track’ of children’s learning throughout year (i.e., formative) and for ‘assessing’ growth at the end of year (i.e., summative).

Portfolios were also considered by educators to be “*useful for showing how children are involved in documenting their own learning*”. For example, one educator commented:

The portfolio explains how the children have been involved in their own documentation and how the observations have all come together.

Likewise, several educators and CDs noted that portfolios are useful because they are accessible to children and families and are useful for sharing information with families and showing children’s progress.

However, one CD commented that the portfolio was particularly time consuming:

Don't have time – a portfolio is the collation to show children's distance travelled. Previously they did do reports but teachers were doing this in their own time.

An educational leader also commented on the time required to go through the portfolios at the end of the year to prepare their transition to school summary statements.

FLOOR BOOKS

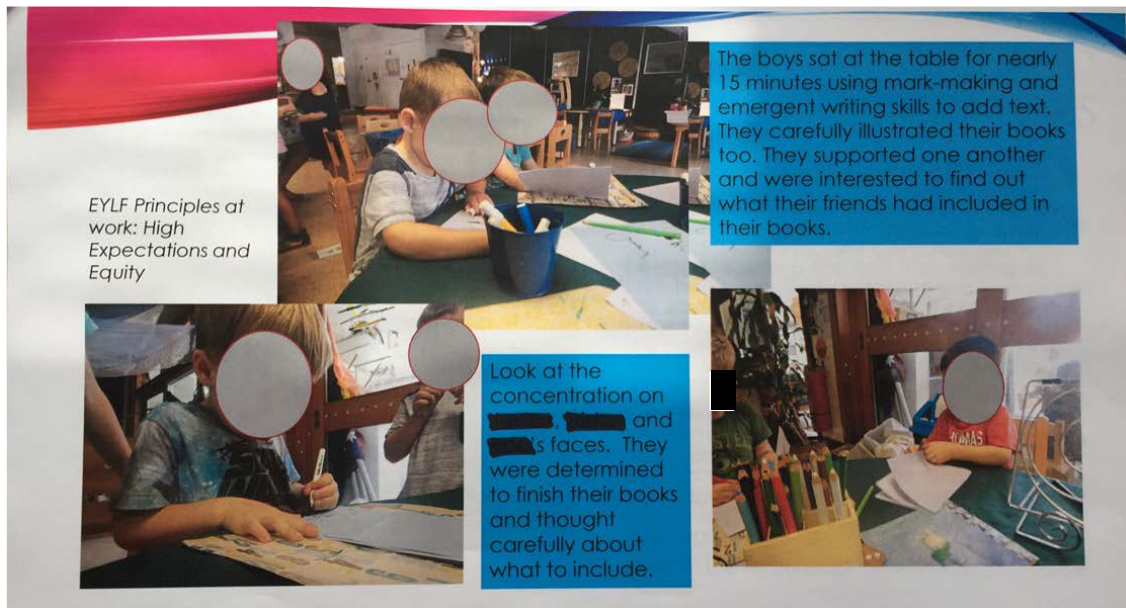
Floor books (written about by Claire Warden, www.claire-warden.com) were used by 15 participants. Of these, 14 provided additional ratings, which were very high on suitability for involving children in the assessment and sharing with other educators (Means = 4.6) and high on suitability for children with a range of abilities or children who do not have English as their home language (Means = 4.4). Ratings were lower for usefulness for assessing children’s learning in relation to the EYLF Learning Outcomes, and for providing ongoing feedback, reflection and planning for individual children (Means = 3.6). Floor books were also rated poorly for informing the writing of a Transition to School statement (M = 3.1) (see Appendix E).

A few educators described their use of floor books; for example, one educator explained that:

For planned activities, she takes photos, writes observations, then writes a summary of what happened in an accessible form for families, prints and sticks these in a Floor Book. She talks about the experience/photos with the children, who are invited to add to the book ... The floor book is also for parents to know about the cultural work that is being done in the centre.

Floor books were reported as being easy to use and highly individualised. For example, one educator commented that floor books:

Are organic, there are no rules to how they are to be used. Everyone can use them in a different way.

Figure 4: Example of Floor Book

A strong benefit of floor books noted by some educators was that they can be used to involve children – “Sometimes with the whole group or a small group of interested children” - in documenting their own learning. One educator explained how in his/her service floor books promote child engagement:

Photos are taken during the day - children pick which photos to go in the floor book. They retell the story. Children can draw pictures in the book.

In this way, some educators considered that children are able to reflect on their own learning from floor books and to contribute to planning follow-up activities. One educator noted, for example:

Children get more opportunity to reflect on their own learning and what happened during the day - they have a voice in what is put in the book. It opens up conversations about their day and what they liked about the day, and what they might like to do next.

However, educators also noted that there can be limits on children’s engagement. As one educator commented, for example, the use of floor books “doesn’t really engage everyone, but it’s still useful for imaginative thinking for children”. Another challenge with floor books is that they can be time consuming to complete. And there was a comment from one educator that parents prefer on-line postings to floor books.

PHOTOGRAPHS

About half of the participants indicated that they used photographs (alone) as a formative assessment tool, of whom 9 completed additional ratings. Very high scores (Means = 4.9) were given for ease of use, suitability for children with a range of abilities or children who do not have English as their home language, informing the writing of a Transition to School statement, and usefulness for assessing children’s learning in relation to the EYLF Learning Outcomes. However, photographs were rated poorly for time to complete the assessment (M = 2.4) and were not strongly endorsed in relation to usefulness for providing ongoing feedback, reflection and planning for individual children (M = 3.6).

Educators said that they liked the immediacy and spontaneity of photographs. They noted that photographs are objective and accurate and often capture what words can’t. However, some educators commented that they weren’t always quick enough to capture the learning moment in a photograph.

The use of photos was viewed by educators and CDs as a potentially good approach for including children in the documentation of their learning. They noted that children can take photos of their learning so that their voice is heard; and photos enable children to look back on their growth.

Educators also commented that photographs can be useful for tracking children's growth over time. In particular, it was noted that capturing a series of photos can document evidence that educators can reflect back on later, to enable them to follow a child's progress in a particular learning focus (e.g. literacy). For example, one educator mentioned that she/he takes photographs of samples of children's writing and notes their progress over time.

Photos were also seen as good for sharing children's learning with families. For example, they can be used in parent teacher interviews and photos can be used to communicate about the child's day and potentially facilitate communication. As one educator explained:

A photo slideshow might lead to conversations with parents – for example, what was the child doing and how can they continue.

Further, some educators and CDs noted that photos of children in their home environment can be used to support the home / service connection.

Nevertheless, some educators expressed concerns about taking photographs. For example, concerns were raised that taking photographs (i) may interrupt children's play; (ii) photographs may be used to replace gathering of children's work; and that (iii) not all children want their photograph on social media. For example, one educator commented that:

Children like to make their work seen and shown rather than the photos.

Significantly, one educator commented that photos are not useful unless educators engage with and reflect on the photos. Other educators took a pragmatic approach noting that photographs should not be the sole use of documenting children's learning – but that they do have a place. In addition, the high financial cost of printing photos was mentioned.

CHECKLISTS

Several educators and CDs noted that checklists 'have their place' particularly to assess 'basics'. As one CD said: "*a long narrative is not always necessary*". Educators mentioned that they completed checklists as children engaged in activities and that they are generally done over a few days.

The majority of participants (19 out of 27) said they used checklists, and 18 provided additional ratings for this type of assessment tool. Ratings were mixed: about half of these criteria received scores of 4 or above, and the others received ratings of 3. Items rated lower than $M = 3.5$ were: suitability for children who do not have English as their home language, suitability for involving children in the assessment, and the time needed to complete the assessment. The highest ratings were given for ease of use ($M = 4.6$) and usefulness for providing ongoing feedback, reflection and planning for individual children ($M = 4.5$).

The types of checklists mentioned by educators included:

- (i) developmental checklists (i.e. checking for developmental 'milestones' and skills such as scissor grip, cutting skills, gross motor) including the *Ages and Stages Questionnaire*.
- (ii) 'found' or service-created checklists (e.g. based on EYLF learning outcomes);
- (iii) checklists focussed on specific curricula areas such as numeracy and literacy; and
- (iv) checklists to test children's knowledge of colour, number and shapes. For example: one educator noted that she/he:
Downloaded a checklist re: colours, shapes, number. I go through and think about what each child needs to work on.

Many educators reported using checklists that they had created themselves. For instance, in one service educators had designed a 3-page assessment that they used to record children's understandings of shapes (circle, square, triangle), and numbers from 1 thru 20. In another, the CD and staff had developed a 2-page comprehensive checklist that included counting, colour recognition, writing skills, cutting skills, and other

areas that supported the EYLF Learning Outcomes. All educators were expected to use the checklist across this large service.

Educators in another centre had developed their own developmental checklist based on the EYLF and Starting Blocks developmental milestones (<https://www.startingblocks.gov.au/your-childs-development/>). They had found this checklist beneficial and the information gathered from it informs and permeates all their practices. This checklist lists developmental milestones expected at particular ages, and children are rated as having 'no difficulty', 'little difficulty', 'much difficulty', or 'milestone not yet met'. The tool is completed by the educators at the start of term and used to set the goals and to think about the activities children will engage in to meet those goals (based on the interests of the child). The completed checklist is shared with parents. The tool has been used consistently in the centre for many years. In this service, in particular, educators noted that:

Previously we relied on learning stories, but we found them too time consuming. It's more important to be with the children rather than be stressed with work. Now we use an in-house developmental checklist to aid in identifying learn goals for the term. Once learning goals are established, we plan activities to align with those goals. We make short jottings for each child in the activities. Once a child has achieved the learning goal, a longer educational summary is written based on the jottings and collected photographs.

This sequence is illustrated in Figures 5a, 5b and 5c.

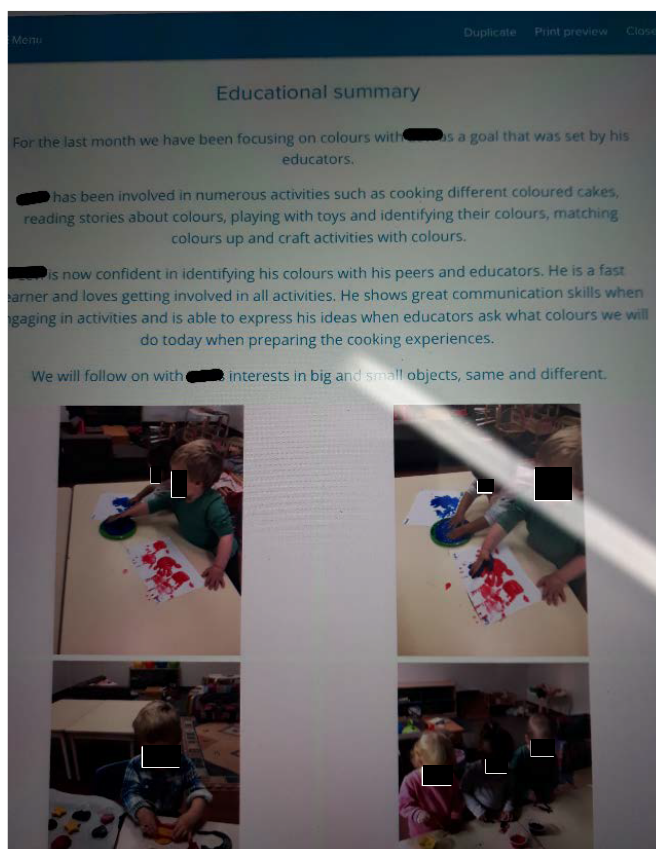
Figure 5a: Example of Service-Designed Checklist

Development checklist				
Child's Name [REDACTED]				
Milestone	No difficulty	Little difficulty	Much difficulty	Milestone not yet met
Social skill				
Maintains eye contact when talking	✓			
Complies with request: Packs away		✓		
Is able to manage emotions			✓	
Waits patiently for adults attention		✓		
Can focus at a table or desk for 10 minutes	✓			
Separates from parents easily	✓			
Can independently begin activities	✓			
Asks for assistance and accepts assistance	✓			
Holds focus for 20 minutes or more	✓			
Plays co-operatively with friends for 20 minutes	✓			
Is comfortable in sharing toys	✓			
Is able to initiate turn taking in a group	✓			
Shows empathy and understanding		✓		
Independent skills				
Puts on and removes clothing	✓			
Can manage shoes with buckles/Velcro				✓
Can manage large buttons on clothing				✓
Goes to the toilet independently	✓			
Recognizes the need to blow nose		✓		
Can manage lunchbox, unwrap food	✓			
Uses spoon and fork independently	✓			
Unscrews drink bottle	✓	✓		
recognise and takes care of belongings	✓			
Shows cleanliness; packs away		✓		
Enjoys Storytime		✓		
Communication skills				
Initiates conversations	✓			
Speaks clearly while maintaining eye contact	✓			
Speaks neutrally; not shouting or whispering		✓		
Answers 'who', 'what', 'where' and why questions	✓			
Is able to recount experiences or stories	✓			
Recites rhymes and songs	✓			

Figure 5b: Jottings

Child	Goal	Jotting	Jotting	Jotting	Achieved
[Redacted]	Shape recognition colours	11/6/19 - recognises Square, Triangle & Circle. 1/4 recognises Circle, triangle, heart, star	12/6/19 counts to 10 by herself. recognised pink yellow 29/8 orange 25/8	recognised Circle 29/8 counts to 5 by herself 25/8	
[Redacted]	Shape recognition	11/6/19 - recognises Square, circle, triangle and heart.	recognised red + yellow 16/9 16/9	counts to 10 16/9	
[Redacted]	colour recognition same/different	12/6/19 recognises pink, blue & yellow	11/7/19 Matching coloured cars to colours in book e.g. red/blue/green		colour recognition achieved
[Redacted]	Regulating emotions shape recognition	5/6 - recognises square, triangle, circle heart, net, hexagon. 11/6 recognises heart & square + star	25/6 imogen was able to count to 10. she skipped six and said 7, 8, 9	recognised Triangle 29/8 counts to 10 16/9	

Figure 5c: Educational Summary



A particular benefit that educators noted for service-designed checklists is that they can be constantly improved and revised.

Several educators and CDs alike noted that the evidence gained from checklists can be used to plan programs and also to demonstrate children's growth and development, through testing at one point and then retesting at a later point. For example, in one service:

Information on the whole is gathered. It is across all areas and the teacher can see where they are sitting in regard to development and achievements. We can see which areas they need support - which feeds into the intentional teaching part of the program.

Checklists were considered by some educators to be particularly useful in the child's year before school, for planning, showing children's progress (e.g. beginning, mid-year, and end of year testing), communicating with other professionals and funding bodies, and in supporting decision making about transition to school. For example, one educator mentioned how the school readiness checklist helps them make decisions about children's transition to school.

We do it at the beginning of the year and then at end of year to see if the child has met the expectations. It helps us decide whether the child is ready for K next year.

Figure 6: Example of School Readiness Checklist

School readiness checklist

Social skills:

- Maintains eye contact when talking
- Complies with requests, example; packs away when instructed
- Is able to manage emotions
- Waits patiently for adult attention for several minutes
- Can focus on a table task for ten minutes, with assistance
- Separates from parents easily
- Can independently begin activities
- Asks for and accepts assistance when necessary
- Holds focus for 20 minutes or more during play-based activities
- Plays co-operatively with friends for 20 minutes or more
- Is comfortable in sharing toys
- Is able to initiate turn-taking in a group
- Shows empathy and understanding towards others

Independence skills:

- Puts on/removes socks, shoes, jacket etc.
- Can manage shoes with buckles or Velcro
- Can manage large buttons on clothing
- Goes to the toilet independently, including self-hygiene
- Recognises the need to blow nose and effectively wipes
- Can manage lunchbox; unwrap sandwich, ensure enough food for afternoon tea, place rubbish in the bin etc.
- Uses a spoon and fork independently
- Unscrews drink bottle
- Recognises and takes care of belongings
- Shows cleanliness; packs away and acknowledges mess

Communication skills:

- Initiates conversations with both children and adults
- Speaks clearly and audibly while maintaining eye contact
- Speaks neutrally; not shouting or whispering
- Answers 'who', 'what', 'when', 'where' and 'why' questions
- Is able to recount experiences or stories
- Recites rhymes and songs
- Comprehends and responds to requests
- Can complete 3 requests in sequence

Checklists were considered by educators to provide 'hard' / objective evidence and consistent information. They could be used to confirm educators' thoughts about a child's development and identify their areas of strength and need. Checklists were considered by some educators to be particularly useful for novice staff. The 'objectivity' of information gathered from checklists was considered to be particularly useful for facilitating 'difficult' conversations with families.

On the other hand, educators also noted some limitations of checklists, including that they do not include children's interests; may only provide a snapshot on any given day and so might not be true representation of

a child's ability; can be time consuming; and may not always be holistic. However, given the contrary views illustrated above, limitations depend on the nature of the checklist.

There were also divergent views on the use of checklists for children with special needs. Whilst some respondents considered checklists to be particularly useful for children with special needs, others thought that they "*might not be the most suitable for children with developmental needs*".

CHILD COMPLETED ASSESSMENT TOOLS

Over one-third of participants spoke about their use of child-completed assessment tools; ten provided additional ratings. All ratings were over 4 out of 5, with the highest being given for ease of use (M = 4.8), suitability for sharing with families and with other educators (Means = 4.9), suitability for involving children in the assessment (M = 4.9), and usefulness for providing ongoing feedback, reflection and planning for individual children (M = 4.7).

Educators described using this form of assessment for skills related tasks, such as drawing shapes, or cutting shapes and pasting them onto a similar shape. An educator explained being "*able to see where they are at, at that stage and what you may need to work on such as scissor grip*". Child-completed assessments were sometimes included as part of a checklist assessment, as illustrated in Figures 9 and 10.

Another structured example was drawing

Each child has a drawing book. An exercise book... generally use lead pencils - may use coloured at times. They are available to the children, but they don't tend to use them themselves. It may be structured drawing - following the teacher's example on the smartboard. Not always as structured - children may do free drawing. Parents may be involved at the pick-up time if children are drawing.

Other types of child completed assessment also provided structure, but "*not limited by adult vision of what the child should be learning.*" The child has freedom to choose, as an educator explained; these were:

driven by the child's interests, family or other relationships. They choose to do it. There can be some structure (e.g., I can write my name. I am this old. I am this tall. This is me. This is my family.)

Figure 7: Example of Child-completed Assessment

Date: october 2019

I can write my name.
MC

I am this old.
5

I am this tall.

This is me.

This is my family.
 This is my mum
 This is my dad
 this is my brother

Figure 8: Example of Child-completed Assessment as part of an educator-completed checklist

Name: _____ Date: 15/10/19

I know these colours.

Brown

I know these shapes:

I can recognise the following numerals.

3 8 9 5 2
7 4 1 6 0

I can rote count up to 7.

I can count 1:1 up to 12.

Child's name: _____

Shape recognition:
 What does the child call these shapes?

Colour recognition:
 What does the child call these colours?

Ask the child to count to 10 or 20 (developmentally appropriate)
 Brief description of the child counting:

Number recognition
 Circle the numbers the child is confident with if developmentally appropriate.

1 2 3 4 5 6 7 8 9 10
 11 12 13 14 15 16 17 18 19 20

FAMILY COMPLETED ASSESSMENT TOOLS

Participants agreed that family input was key to formative assessment. About half used this type of assessment, and 13 participants provided additional ratings. The highest rating was given for suitability for sharing with families ($M = 4.8$), but educators were less confident about suitability for children who do not have English as their home language ($M = 3.5$). Lower ratings were also given for usefulness for informing the writing of a Transition to School statement ($M = 3.3$), assessing children's learning in relation to the EYLF Learning Outcomes ($M = 3.4$) and the time needed to complete the assessment ($M = 3.9$).

Many of the examples were 'All about me' (or similar) forms (see Figures 11 and 12) that were completed by families at the beginning of the year and were seen as useful for facilitating family input. These forms gather information from the family about the child's interests, strengths, "*learning development, communication level, abilities and interest*" and what is important to family (e.g. goals and aspirations for their child). Educators agreed that this information adds to their planning and can alert teachers to new information. As one educator explained:

At the beginning of the year, the parent fills out the form. It's about the child's interests, abilities, family background, heritage. It's called "all about me".

Similarly:

At the beginning of the year 'All about me' forms are used to create goals for the children as a group. Individual goals are written down and common goals are condensed into themes. This allows educator to go to back to the goals and connect to what the parents want. Interviews with families can also alert teachers about what they need to pay attention to (e.g. if parent comments that a child does not want to come to school – why is that?)

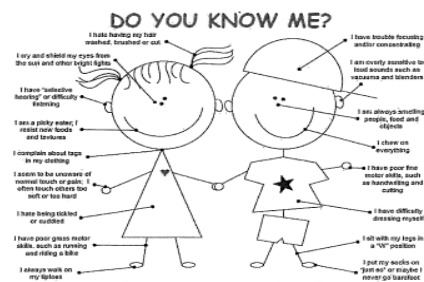
Several educators, lamented, however, that getting family feedback is not always easy: "*When we send it out to the families, it's hard to get them back.*" Also mentioned, was the difficulty in completing these forms for families who had limited English.

Figure 9: Example of Family-completed Assessment

Child Name: [REDACTED]		Preschool 2019		Date:	
<p>BEING</p> <p>What does your child enjoy about preschool? <i>He enjoys being social with the adults and the older (big) kids.</i></p> <p>What do you like to do at preschool? <i>Play with the bikes</i></p>			<p>BECOMING</p> <p>What would you like us to focus on for your child in preschool? <i>Name writing → Being school ready.</i></p> <p>What would you like to learn? <i>My rules</i></p>		
<p>Hobbies & interests</p> <p>How does your child like to spend their time at home or away from preschool? <i>Likes the ipad and enjoys home dancing ("we will rock you")</i></p> <p>What do you like to do away from preschool? <i>Go to the park.</i></p>			<p>BELONGING</p> <p>Where does your child belong? (cultural group, family groups, neighbourhood group) <i>He belongs to our family and the neighbourhood.</i></p> <p>Where do you belong? <i>At [REDACTED]</i></p>		
Date:	Parent/carer:	Child:	Teacher:	[REDACTED]	[REDACTED]
Review Date:	Parent/carer:	Child:	Teacher:	[REDACTED]	[REDACTED]

Figure 10: Example of Family-completed Assessment Tool

8. Does your child have any favourite toys, games & activities?
Favourite toys are dinosaurs which he calls dinorahs and does not like to share them !!
9. What are your child's interests & hobbies? Including after preschool activities e.g. soccer, football, dancing, Little Athletics, Pony Club
does not have any afterschool activities - Loves reading Books and painting, playing with playdoh.
10. Does your child have any fears or traumatic experiences we need to know about? e.g. storms, spiders.
During storms - Does not like thunder.
11. What calms or settles your child if they are upset?
distraction - usually a hug - Talking calmly.
12. What are your child's strengths?
independance - like's to attempt most tasks on his own
13. Are there any areas of your child's development that concerns you? e.g. social, physical, speech, behaviour. Please see "Do You Know Me?" diagram below
difficulty with speech - Social (does not like to play with other kids much).
14. Is there any other relevant information that may benefit your child's experience at _____ Preschool?



DIGITAL APPS

Educators, CDs and APs alike endorsed the use of digital platforms, but feelings were mixed. A large number of participants (22) provided additional ratings that, overall, fell between 'neutral' (lowest rating was 3.2) and 'good' (highest rating was 4.44).

Educators noted that digital apps are flexible and creative, but the structure also assists educators to write their assessments of children's learning. As one educator noted the digital app is:

Flexible - but also guides the educators ...it is creative ... we get to use videos and photos to document learning, which is very useful.

Comments made by educators and CDs noted that, in general, digital platforms are particularly useful for sharing photos and information, and communicating with parents, as parents can access it and make comments. Likewise, digital apps were considered by APs to be versatile and accessible throughout an entire service and can provide a central database of information.

Nevertheless, concerns were raised about digital apps. Further, some CDs were concerned that the centre does not own the information. Additionally, APs were concerned that there are potential risk issues regarding un-authorised access to information when iPads and other digital technologies are used. An AP raised concern with the limited degree to which children can be engaged in the use of digital apps. Some educators noted that it was difficult to engage or reach families using digital apps. They expressed concern that families

may have limited knowledge or online access and that families may have difficulty navigating systems. Some educators noted that despite the approved provider holding sessions to teach families about the app, there needed to be more education on how to use it (for both educators and families). Another educator commented, for example, that digital apps are “good if educators are time poor, but they cannot replace other more effective and deeper ways of assessing”.

The digital platforms specifically mentioned by educators included:

Kindy Hub: Kindy Hub has features that educators found useful, such as voice and video recording which are able to record how children interact with one another. It was considered by some educators as having a useful template for guiding the educator and supporting critical reflection.

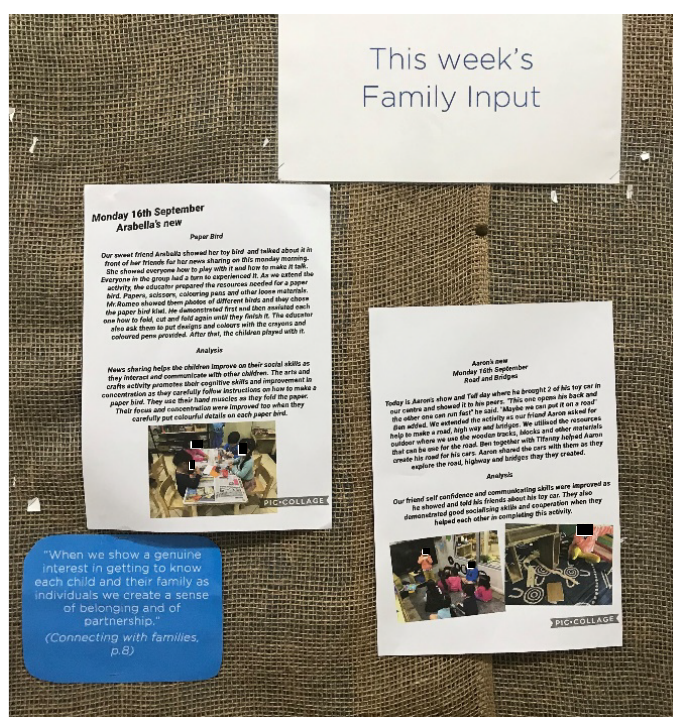
The template guides the educator. It links to goals, theories, observations and reflections for educators. It asks questions like, "how did you support the child's learning?", and encourages educators to reflect.

Pencil or PicCollage: This was used by educators as a tool for recording observations and progress, across all EYLF learning outcomes. It was noted as being colourful but time consuming. And some frustrations were voiced (i.e., Pencil has maximum characters).

Kinderloop: One educator noted that Kinderloop was useful for providing kindergarten teachers with “insight to the child before they start (school)”.

Storypark: Storypark was used by some educators as a daily diary. In some services, educators reported, Storypark was considered a comprehensive tool. It was used to record children’s individual goals, and it was also seen to assist with the development of transition to school statements. There were some concerns raised with Storypark by educators, including that it was time consuming and takes educators away from children; and that the layout is problematic.

Figure 11: Daily Journal - shared on Digital App



INDIVIDUALISED / PERSONAL LEARNING PLANS

A very few educators mentioned individualised or personalised learning plans. Examples were provided of individual education plans (IEP) that included child’s interests, strengths, what children are working towards, and parent goals. One educator mentioned the use of externally provided early intervention plans – for children with additional needs.

DAILY JOURNALS

Daily journals that record children's daily routines and experiences were seen by a few educators as a formative assessment tool. One educator noted that daily journals can be used to record children's interests and ideas.

COLLECTION OF CHILDREN'S WORK SAMPLES

A few educators reported that they collected children's work samples at various points as a 'point of reference'. By comparing these over time, they could observe change that informs their individualised program planning. For example, an educator explained how each child had a 'scrapbook' that was a place for them to use literacy skills, write and draw, with educator support. These were introduced as part of the Pre-Lit program. Educators read classic stories, speak about the authors and illustrators, and encourage children to tell a story.

ON-LINE ASSESSMENT TOOLS

On-line assessment tools were endorsed by a small number of participants, with six providing additional ratings. The highest ratings were given for usefulness for providing ongoing feedback, reflection and planning for individual children ($M = 4.7$), and suitability for children with a range of abilities ($M = 4.5$). Other ratings ranged from 3.2 to 4.3

A couple of educators noted the use of on-line assessment tools and that these were easy to use and good way of sharing information with families and tracking children's learning. On the other hand, they were also considered to be time intensive, not helpful for parents, and some educators considered that educators know more than the tool shows.

TRANSITION OF ROOM ASSESSMENT

In one service, children were formally assessed prior to moving from one room to another – to support their smooth transition. An educator in this service stated:

It is important at the beginning to know the child's development. It helps them transition between rooms and means the teacher can follow their interest and set up experiences catered to them.

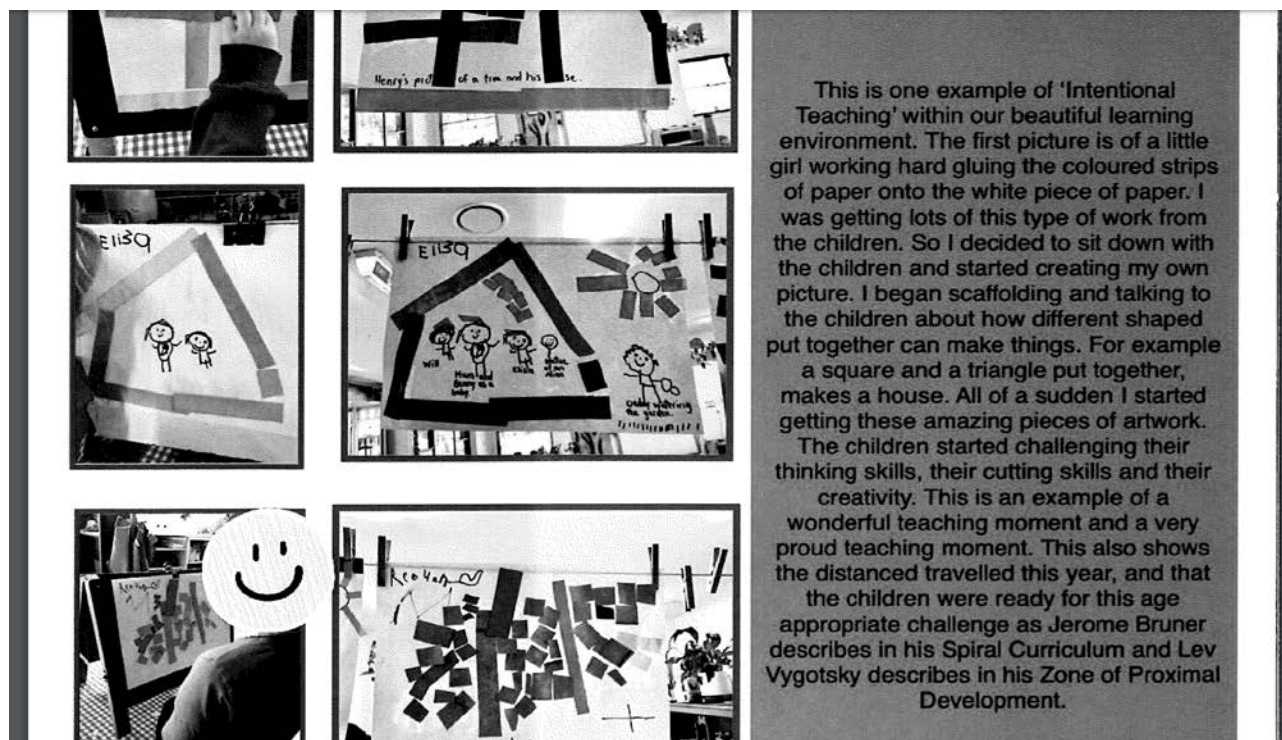
The same educator commented that the tool is also useful for highlighting the children's change over time as they move across rooms and this is particularly rewarding for educators:

It's great seeing the development of children and how they change over time. - great to look at the children interest and where to go from there.

TEACHER REFLECTION

Interestingly, many CDs spoke about the importance of reflection as a formative assessment tool. In some cases, teachers' reflections on a learning experience were included in a piece of documentation shared with parents (see Figure 12). Reflection was noted as being done 'formally' through daily diaries or reflection books, as well as in team meetings, but most often 'informally' as educators went about their daily work. For example, an educator described how she would think about something she had observed, then write about it in the reflective journal and pose questions about a child's activities and how to interpret this in the centre's reflective journal, and invite other educators to write their thoughts. These 'sharing places', created spaces where all educators could communicate with each other and plan 'next steps' together.

Figure 12: Example of Teacher Reflection on a Learning Experience



PRACTICE MODELS FOR INTEGRATING EFFECTIVE FORMATIVE ASSESSMENT TOOLS

When asked about their practice models for implementing FA, respondents' comments coalesced around six main themes:

- (i) the need for FA to link to the EYLF and NQS;
- (ii) the role of FA in the planning cycle;
- (iii) the utilisation of a focus child system to support FA;
- (iv) the importance of FA being responsive and inclusive of children's voices;
- (v) philosophical perspectives on the need for FA to be strengths-based and holistic; and
- (vi) the need for FA to be inclusive of and accessible to families.

FA NEEDS TO LINK TO EYLF AND NQS

There was strong evidence from APs and CDs of the need for FA to link to the EYLF and NQS. Interviews with the CDs strongly indicated that the impetus for developing FA tools and procedures in their services was to meet NQS. The following is an example of one service's journey that took a strategic approach:

The programming and planning cycle was changed with the new NQS. Five to six staff did a professional development course in 2015 for 5 weeks (from CELA). From the training, they implemented a new assessment system in 2016. Staff can do it their way, but within the guidelines.

In 2018, we had to make sure that they were operating within the new NQS guidelines. I went through the new NQS and copied and pasted what the assessors would cite. This document informed the planning cycle. We made sure that the cycle fit in with what the assessors were looking for. The NQS asks for consistency of practice – this is achieved by having an over-arching planning cycle. The original diagram we use came from the EYLF Guide. In 2018 it was personalised for the centre. The child is always in the centre of the program'.

The following are further examples of the journey services took to address formative assessment in relation to NQS ratings. The Centre Directors (CD) reported:

The centre was using a digital app to document and reflect on children's learning, but there was no "assessment - no in-depth part to it". Staff were using a variety of methods and were not working as a team. It was not inclusive of all staff. They needed consistency, for the staff and the children. The planning cycle needed attention. The CD introduced a new format for recording observations, analysis of learning, planning and reflection that was easier for the staff to use, share and talk about. There was some resistance to the change, "being told what to do".

I have been here three years. I identified a good depth of knowledge and passion amongst the team, they are connected to the community. However, I noticed they were changing things daily e.g. Lego one day, mobile the next. I felt overwhelmed and wondered how would the children feel? We investigated the continuity of learning for the children. A lot of talking about the program - the planning cycle wasn't so evident. We have many children in the service so can't work on every child's interests. Intentional teaching – we make notes in a book and will follow up later. Opportunity for children to go back and master the skills. Used ACECQA folder - The Guide to NQF - to support the approach. A strengths-based approach that makes the planning cycle more visible and addressed inconsistencies across both rooms. Now feel we are in a good place.

Correspondingly, many ECTs, ELs, and Eds spoke about the importance of being able to easily link EYLF outcomes to FA tools. And there was strong evidence that educators' use of formative assessment tools was linked to the EYLF, including the use of the NQS as a point of reference and the inclusion of EYLF outcomes on FA tools. Indeed, many educators made comments in their assessments of children's learning similar to the following: "based on expectations...and aligned with the EYLF learning outcome".

PROGRAMMING AND PLANNING

Most CDs clearly linked the use of FA tools to the planning cycle of observe, critically reflect and interpret, intervene, observe and onwards. Correspondingly, there was evidence that some (but not all) educators were following a planning cycle. Participants who spoke about using a planning cycle linked different types of formative assessment such as observations, learning stories and reflections together as an informative planning process for educators. An example of this process was described and illustrated in Figures 5a, 5b and 5c, above. Some educators noted the importance of the 'where to next' aspect of the planning cycle.

FOCUS CHILD SYSTEM

Some educators mentioned that they utilised a focus child system to enable them to complete FA. Educators and AP both spoke of the importance of group as well as individual focused observations. Those who used it said that a focus child system was useful because it helped to spread the load amongst educators. One educator noted for example that a focus child system is "Good for our workload because we can focus on a smaller number of children and target their needs."

RESPONSIVE AND INCLUSIVE OF CHILDREN'S ENGAGEMENT AND VOICES

CDs and educators noted that FA and program planning needs to be responsive to, and inclusive of, children's engagement and voices. Whilst some educators discussed how they included children's voices in the FA, for example through the use of mind maps, as described below:

The mind map is the children's voice in our program, what the children have done and that informs our program. We program weekly but stay flexible to give children time to revisit and redo. We stay sensitive on how we move forward to ensure that the children are still interested.

Participants described a number of ways that children are engaged in formative assessment tasks, particularly floor books and child-completed assessments (as illustrated above), and gave moderate to high ratings (Ms = 3.2 to 4.9) for 'suitability for involving children in the assessment process' (see Appendix E) but these tended to be seen as "part of normal play and interactions with educators". Participants referred

to programming for children's interests, e.g., "*we tell them what we are doing today, from the day before, ask them what they want to change*" and children's contributions to assessment records, e.g. "*something to go in the portfolio, taking photos*", but further evidence for the inclusion of children's voices in FA in practice was somewhat limited.

STRENGTHS BASED AND HOLISTIC

Some CDs spoke of the importance of formative assessment being both holistic and strengths-based, within the context of play-based learning. Indeed, there was ample evidence of educators' focus on children's strengths and interests in a range of developmental and learning domains. For example:

Planning for individuals or the group, depends on what we observe. The focus is on children's strengths, not weaknesses. Planning is also based on what children ask for - they work with the children, discuss and negotiate, to plan.

Likewise, there were several comments, such as the following, about FA being 'holistic'.

This holistic approach that we have is very important: we don't want to lose that. We want to maintain the holistic relationship-based approach. It is genuine and real. We don't want just a "tick box" or "structured" and standard tool.

However, it was impossible within the scope of the current project to assess the comprehensiveness of the FA being conducted to determine whether it was indeed a completely holistic approach.

INCLUSIVE OF AND INVOLVING FAMILIES

Participants reported that a primary function of FA is to communicate the child's learning and development to families, and facilitation of communication with families was important in their choice of FA tools. This was largely reported as two-way communication with families, where not only did families receive information, but they could also provide input into the assessments. For example:

Parents are asked to provide a set of goals for the child at the start of the year. Day to day we learn from families what the children are interested in and what activities they like. Sometime children bring activities in from home and we use them with other children. Some parents provide feedback via Storypark.

Consequently, educators and CDs alike noted the importance of formative assessment tools and processes being accessible for families. Educators attempted to make information accessible and available for families and noted that what works is different for different families. As one educator commented:

I believe formative assessments should be informal, clear and easy to understand and communicate with families of the children.

However, it was noted that some families' capacity to access FA or provide input is limited. Whilst participants gave relatively high ratings (Ms = 3.8 to 5) for 'sharing with families' for the FA tools they used (see Appendix E), the degree to which FA was accessed by families in the services visited was unable to be determined.

EVALUATION OF THE FA TOOLS BEING USED IN THE SERVICES

VALIDITY OF FA TOOLS

Participants were asked their reasons for choosing a particular FA tool and whether any evaluation of their FA tool and practice had been conducted. There was no evidence in the AP, CD, ECT, EL or Ed interview data that the FA tools being used by the case study services have been evaluated for their psychometric properties, e.g., their validity or reliability, or that they had been selected on the basis of research evidence. Rather, the responses indicated that specific tools had been selected based on suggestions from colleagues or ease of use.

When FA tools were considered by respondents to have been ‘evaluated’, these ‘evaluations’ had been conducted by informal internal review or external review by a consultant, the NQS A & R assessor or spot checks by the Department of Education. Most responses referred to an external review of FA through the NQS Assessment and Rating process. Some educators talked about developing FA processes in line with expectations of A & R requirements and assessors’ feedback. There is no mention of FA development being part of the services’ Quality Improvement Plan. A number of CDs and educators noted that their FA tools had been subject to review by an external consultant, sometimes in the form of a ‘mock’ assessment and rating visit. For example, one CD mentioned that:

Two consultants came in to do a "mock" A & R. They provided feedback to educators about formative assessment.

Some CDs noted the high costs involved in having consultants conduct these mock A & R. Others reported that review of FA tools and procedures was primarily conducted internally, by educators and program leaders, and sometimes involved families. Participants commented that the development of FA tools was an ongoing process based on trial and error, with the flexibility to make changes according to the needs of children, educators and best practice.

IMPROVED TEACHING PRACTICE

Approved Providers, CD and educators alike recognised the importance of FA and its potential for changing practices. There was some agreement amongst AP that formative assessment tools had prompted critical reflection on practice, although they could point to no clear evidence that the reflection had resulted in actual changes being made to practice. CD and educators, on the other hand, provided clear examples of where the use of FA had resulted in shifts / adaptations in educators’ planning, and practice / teaching strategies.

Formative assessment informs planning.

The FA tools that were rated very highly (Ms > 4.7) by participants as being useful for providing ongoing feedback, reflection and planning were observations, learning stories, service-designed tools, child-completed assessment, on-line assessment tools, and externally-sources assessment tools (see Appendix E). There was strong evidence that FA informs educators’ planning, including their choices of teaching strategies. Educators used FA to gather information on children’s strengths and interests and plan accordingly. For example, one SSLO said she

Uses children's interests or questions to design an activity for the group. She finds a way to get to children's understandings of culture through interests for example interest in bugs and insects as an experience to get at a child's sense of country.

Whilst it was clear that FA informs educators’ planning, some educators commented that assessment documentation was time consuming and that less documentation allows them to spend more time with children.

FA leads to changes in educator assessment practices.

Participants were also able to demonstrate how they had made changes in their assessment practices – some seemingly positive changes, others neutral and some potentially troubling. For instance, there were shifts reported from educators’ use of domain specific checklist assessments to holistic assessments; from a focus on group assessment to individual assessment; and a change in the amount of child involvement in FA - with some moving to more involvement and others moving to less involvement of children. The comment below demonstrates one CD’s reflection on a shift in FA practices from checklist to holistic assessment following professional development:

Four years ago they were doing checklists and observations against outcomes they wanted children to achieve. It was very outcomes based and they were not looking at the whole child. We are attached to a school and it was more about what was needed to get a child ready for school. We received training in line with play based learning. For me this was more beneficial than looking at checklists. Doing assessment in the context of play, we get a better picture of the child. It's about balancing what's important for the school vs what is important for families at home. Families are very verbal and communicating about assessments is very important - having conversations with

families about what they want their children to do. We look at the EYLF and think about what the child needs as a whole, not just in terms of school readiness.

IMPROVED CHILD OUTCOMES

Participants were asked which of the FA tools they thought was most effective in showing change or progression in learning. Responses varied, but most participants referred to some form of checklist or service-designed assessment tool that was completed with the child or families 2 or 3 times per year. The information was used to feed into summative assessment reports (e.g., school readiness checklist / transition to school statements). The examples we viewed demonstrated learning progression; however, it is impossible to provide definitive evidence that use of FA tools has impacted on child development outcomes.

It was clear that different FA tools were being used by staff to document learning goals, achievements, and planned activities to extend learning; that is, the tools were being used as part of the programming and planning process for individual children. In this respect, we could expect that use of formative assessment will be having a beneficial impact of developmental outcomes; however, with few exceptions, the tools being used did not provide a clear or standardised format for showing evidence of developmental progress across the year before school.

There was a sense, communicated by educators, that work samples allowed “*educators to see changes in the children’s skills and abilities over time*” and could demonstrate children’s progress. Work samples included drawing and writing, in individual children’s collections. One educator reported the service’s use of ‘scrap books’ for literacy, drawing, language and writing that were introduced for the PreLit Program (<https://multilit.com/programs/prelit-program/>). The program was used daily and was felt to be contributing to improvement in children’s literacy skills:

She is using the PreLit program/lesson plans and sees good improvement in the children's literacy skills. It provides evidence of progress across the year.

EFFECTIVENESS OF FA TOOLS FOR STUDENTS WITH ADDITIONAL NEEDS

CDs, ELs, and ECTs said that they receive and respond to requests by external consultants to conduct FA of children with additional needs. In some cases, they were asked to use FA tools provided by the external agency (e.g., Language Scales by the speech pathologist), but in others they use their own checklists or an external screener, such as the Ages & Stages Questionnaire (<https://agesandstages.com/>).

Participants rated most FA tools as being suitable for use with children with a range of abilities (Ms > 4.5). Particular benefits were noted for observations with photographs, photographs alone, learning stories, service-designed tools, on-line assessment tools, and externally sourced assessment tools.

Lower ratings were also given for suitability of FA tools for children who do not have English as their home language; only observations with photographs, and photographs alone received ratings over 4.5 out of 5.

Further to this, there is no evidence in the interview data that could demonstrate the effectiveness of FA for students with additional needs.

MODELS OF PROFESSIONAL DEVELOPMENT ACCESSED BY SERVICES TO SUPPORT FA

Educators reported that they had accessed a range of internal and external professional development opportunities to support their implementation of FA. The most common types of professional development mentioned were:

- In-house mentoring, collegial support, staff meetings and induction
- Networking (in person and on-line)

- Professional development training with external services

IN-HOUSE SUPPORT FROM COLLEAGUES / APPROVED PROVIDER

In-house development and support from staff within the larger organisation were the most common forms of professional development noted by CDs. In-house development included induction, mentoring, and support from educational leaders, centre directors and colleagues. Likewise, many educators spoke of working with colleagues and leaders to develop their knowledge and processes around FA. For some educators, processes were already in place and they learnt 'on the job'. For example, one CD explained the process of inducting new staff into the use of the FA in their service, as follows:

Once they [the educators] start, they are mentored. A full day with the CD before they begin. Then the Ed Leader goes in every week to spend time with them. There are also experienced staff in the rooms. Everything is documented so things are easier to follow. The feedback is that it can be overwhelming. There is a lot of formative assessment here. That can be stressful for staff.

NETWORKING

Networking was also commonly referred to by CDs. Centre Directors spoke about groups of services (usually from the same service provider) working together on FA. Centre Directors reported a range of methods for networking, including face to face meetings and online, for example:

Have done PL with other centres so we see what other practices people are using.

PROFESSIONAL DEVELOPMENT TRAINING

Professional development training was also common – including multiple day / session workshops. Training providers that had been accessed by the case study services included the NQS roadshow, Semann and Slattery, CELA, Gowrie, DoE engagement officers and online training providers. Some educators also spoke about attending conferences, such as ECA conferences, to gather ideas.

MULTIPLE METHODS OF DELIVERY

ECTs, ELs, and Eds noted that different training and support suits different people. For example, they commented that some people are comfortable with online support whilst others are not. Consequently, many educators combined ideas from a range of sources. For example, an educator in one service commented that:

Different training suits different people and appeals for different reasons. We had a consultant come in. All the ECTs went to the ECA conference last year. Sharing things that they find online. Draw on multiple resources that are available...

Similarly, most CDs noted using a mix of professional development methods to provide support for their staff. For example, one CD said:

Everyone is involved in professional learning

We have a community engagement officer who helps and encourages parents to engage with online resources. Other support officers provide informal professional learning. We also get support from Microsoft Team groups.

Professional development was identified by one Approved Provider as a requirement for staff in relation to assessment, and it was noted that staff may be supported to undertake training that is offered in different modes such as those listed above.

SUPPORTS AND BARRIERS TO USING FA TOOLS

SUPPORTS FOR USING FA

Across the interviews the most common supports for FA included:

- Strong and supportive leadership
- Adequate time
- Adequate funding
- Evidence
- Staff commitment, buy in, openness to change and willingness and ability to work with new systems
- Support of management with change
- Ease of use
- Flexibility in using professional judgement in their choice of FA tools
- Community support
- Professional development and external support, i.e., DoE officers
- Internal induction and staff training
- Availability of different models of professional learning.

Strong and supportive leadership was noted as important. A CD, for example, noted the need for leaders to get “*everyone on the same wavelength [and being able to work] one on one*”; develop skills in using FA tools; and be open to making adjustments for the needs of the team.

Likewise, educators said that having the support and leadership of leaders who know what to do was key to the effective use of FA tools and processes. They said that this translates into commitment from staff to complete assessments.

Time was also considered important as was having adequate resources, staff buy-in/commitment and compelling evidence. One director noted, for example, the importance of backing up processes with research

Time - definitely. Resources - we are lucky. Buy - in from colleagues - in the beginning, there was some resistance. Try to have an article or something to back up suggestions of approaches. Try and follow up with research.

Another described the service’s providing of additional staff to support time needed for assessment:

The school funds an extra SSLO to allow teachers time to do formative assessment with a smaller group of children.

Educators also commented that ease of use of the FA was crucial as was having the flexibility to use their professional judgement in using a range of tools:

We get a really rounded picture of a child using lots of different assessment tools. Not just checklists. Keep things flexible and give educators time to really reflect on what they are doing. Provide professional development so we know we are on track.

BARRIERS TO USING FA

The most common barriers to using FA across the interviews were:

- Large numbers of children – which makes it hard to keep up
- Lack of or insufficient access to resources; i.e., computers or digital access for parents
- Space and time constraints
- Changes in staffing
- Having a ‘one size fits all’ approach
- Lack of access to PD or training around FA
- The costs involved, including those associated with transferring to a new system, paying for PD or paying for relief staff.

Educators said that lack of time and large numbers of children was the key barrier to maintaining effective FA tools and processes. They said that a 'one size fits all' approach can also inhibit, as people work better in different ways. For example, some like paper and pencils, others like computers.

SUPPORT NEEDED FOR THE IMPLEMENTATION OF FA

There were few suggestions from participants about support needed by services for the implementation of FA. Those that were mentioned as important for developing effective FA, were the need to:

- Build educator confidence in their ability to conduct FA; and relatedly
- Provide educators with professional development to develop their knowledge.

OVERVIEW OF FINDINGS AND FURTHER RESEARCH AREAS

CURRENT PRACTICES, CONCERNS AND CHALLENGES

The Stakeholder Consultations confirmed the use of the planning cycle in the selected ECE services. All educators used a variety of approaches and tools to collect evidence of and document children's learning, and to use this information to communicate with families. Whilst many were using digital platforms to store and share documentation, the primary source of evidence was educators' written observations, collected 'on the floor' and either written down briefly in a 'back pocket' notepad or remembered and written later in one of the formats used by the service. More formal writing was entered into a digital or paper recording format, with or without accompanying photographs, or with children's work. Reflection on documented experiences to inform and plan individualised learning experiences was done as a separate step, and in some services only by the qualified ECTs and ELs, not by all educators/SSLOs. Similarly, preparing reviews of individual children's learning over a longer period, such as to prepare a summative statement for children making the transition to school in 2020, was the role of the qualified staff.

This process illustrated the practice model that services typically used to integrate formative assessment into the program. The use of written observations (including learning stories), while preferred by participants and rated as useful for reflecting on learning and planning for individual children, was time-consuming, not only for writing the records, but also for reading, reviewing and collating them. We noted that this model of practice provided little evidence that formative assessment contributed to child outcomes. The tools used were mostly chosen to be holistic and part of a continuous process, to be used to develop final summative statements. However, it was difficult to ascertain the extent to which these open-ended tools were used to assess all aspects of the EYLF Learning Outcomes, systematically for each child. This may be why many services also chose to use (or design their own) developmental checklists. The use of a tool where a 'rating' has to be recorded across all learning and development areas forces educators to look holistically at each child.

We further noted that while it was clear from the stakeholder consultations that formative assessment tools prompted critical reflection on practice, and informed educators' planning, there was little or no evidence that the use of FA tools resulted in changes, improvements or adaptations in teaching strategies.

All stakeholder services were using or moving towards selecting and using a digital platform. These were seen to provide an efficient way to manage the large amounts of written and photographic evidence services collect, and to provide ready access to information to prepare summary statements on individual children. Digital platforms also have the potential to show progress in children's learning over time, and to create summative assessment reports. However, it was noted that this process is only effective if the information entered by educators is high quality, useful, and structured in a way that allows progress or changes to be measured and seen.

FURTHER RESEARCH AREAS

This research has clearly tapped into an issue of growing concern and a challenge that the ECE field as a whole will need to face. To address this challenge in ways that will be acceptable and appreciated by the field, future research will need to draw on the principles of implementation science. Redding et al. (2017) describe implementation science as the study of processes and conditions that promote or impede the effective take-up of evidence-based practices in real-world contexts. Implementation success relies on educator buy-in and participation, and a supportive organisational context.

It is clear that services are looking for or designing their own formative assessment tools. In moving forward to support this direction in ECE, the selection of tools should aim to meet criteria for psychometric validity as well as for social validity (Bagnato, Neisworth, & Pretti-Frontczak, 2014). Social validity refers to the tool's ease of use, its accessibility, its acceptability to teachers and families, and its suitability for supporting communication and collaboration across key stakeholders – ECE educators, schools and teachers in the first year of school, and families.

CLARIFICATION OF THE PURPOSE OF ASSESSMENT

The formative assessment practices that have been identified in this research highlight the trade-off between the nature of an assessment tool (e.g., checklist, observation, work sampling) and the tool's reliability and validity. The less standardised the assessment tool, the more difficult it is to obtain a reliable or valid measure of a particular aspect of learning. For example, when teachers are asked to make judgments about a child's performance based on observations made in the classroom, their ratings may be influenced by the time of day, their relationship to the child, and the nature of the task or activity they have observed. The interplay between the assessment tool and its psychometric properties is one major reason why it is necessary to be clear about the intended purpose of an assessment.

In general, if the assessment is intended to make comparisons across children and classrooms, or to make placement decisions about a child, or to evaluate an ECE program, then a sufficiently standardised assessment is necessary. Standardised assessments articulate explicit administration and scoring routines which must be followed. These types of assessment tools ensure that the judgments made about children and programs are accurate and not biased by sentiments of the assessor. To do this, however, as part of ECE practice, using tools administered by ECTs, ELs and Eds, is difficult, as was noted in the review of literature.

This study has highlighted the tension between pedagogical documentation for the purposes of planning for individual and group learning, and documentation for tracking the process and progression of children's learning, that is, for formative assessment.

FURTHER RESEARCH AREAS

It will be important for ECE providers, policy makers and researchers to work together to clarify the purpose(s) of assessment the year before school, and particularly in relation to gathering and providing information that supports children's transition to school.

MEETING A NEED

Despite the preference in ECE for holistic observational methods, the Stakeholder Consultation interviews clearly identified an appetite for information that demonstrates children's progress toward the EYLF Learning Outcomes. Gathering systematic information about learning progression was also seen as a means of confirming the effectiveness of a service's educational program and teaching practices. The case study services that had addressed this need had developed specific purpose assessment tools that were used at two or three points in time (e.g., knowledge of colours, shapes, numbers), or had designed a checklist to record individual learning. Interestingly, despite educators' reservations about developmental checklists many participants confirmed the usefulness of this approach for:

- identifying children with developmental or learning difficulties;
- confirming their perceptions of developmental or learning difficulties;
- facilitating conversations with families about developmental/learning concerns;
- tracking change or progress in children’s development or learning.

While educators are looking for or developing their own instruments to record change over time, it was apparent that there are currently no universally advised formats and no evidence that the formats that are being used have psychometric validity. In other words, although progress may be shown, there is no standardised measure of change or evidence that change is related to the educational program. Related to concerns raised in the previous point, and highlighted in the literature review, some of the participants also pointed out that checklist formats did not capture change effectively or that ratings might vary according to the time of day, and so on.

These concerns highlight the challenge of finding and using a comprehensive checklist format that is well designed, psychometrically valid, and able to be used effectively by educators with a range of qualifications without bias.

FURTHER RESEARCH AREAS

To meet this need, further research is needed to produce or adapt a ‘standard’ developmental checklist that can be used by services to identify the individual competencies and learning needs of each child. In progressing this option, we note the field’s resistance to the term ‘checklist’. This could be addressed by adopting a different terminology. For example, the term, Integrated Check-Ins (ICIs) has been used to describe brief assessments of skill levels across the integrated scope and sequence of the curriculum (Fantusso, Gadsden, & McDermott, 2011).

The skills, attitudes and understandings assessed by an adapted or locally-designed tool should directly map onto NQS standards for assessment in ECE and also link to indicators of development and learning identified in the EYLF. Once available, a reliable and valid tool could help teachers monitor individual children’s progress and potentially also be used to create a classroom profile of individual differences in children’s content knowledge and ability levels to inform planning for teaching and learning.

SELECTING / ADAPTING A FORMATIVE ASSESSMENT TOOL OR TOOLS

The Literature Review identified 22 formative assessment tools that are currently being used by educators in ECE services and schools, internationally. All 22 had supporting research and/or psychometric evidence and a number are readily accessible online. Notably, however, almost all of these tools had been developed for the United States (U.S.), where the approach to early childhood education pedagogy and practice can be very different from Australia and New South Wales. In the U.S., it is not uncommon for the educators to follow a commercial curriculum package that has been developed by external experts. Most of these curricula have a solid basis in research evidence, and formative assessment instruments are provided as an integral part of the package. For example, the Creative Curriculum includes standardised assessment through Teaching Strategies GOLD, and the High Scope curriculum uses the Child Observation Record. These provide well-established examples of a comprehensive model of ECE practice in which formative and summative assessment are integrated.

We also noted a different assessment tool, the Desired Record of Developmental Progression (DRDP) that is not linked to a specific curriculum package, but was developed by a consortium of experts for use across a variety of services and curricula. The appeal of the DRDP is that it: includes 4 to 5 years of background work to develop the scoring system; provides examples to help guide educators’ scoring decisions and identify indicators of emerging and achieved competencies; and includes internal systems to aggregate observational information to show progress or differentiate children’s abilities. The collaborative history of this tool’s design and production suggests that the DRDP developers would be open to further collaboration, potentially, to adapt the instrument for use in Australia.

FURTHER RESEARCH AREAS

There is no currently available tool that can be taken off the shelf and immediately implemented in ECE services in NSW. Further research could be undertaken by forming a collaborative working group (e.g., an independent university partner in collaboration with CESE and the ECED) to conduct a thorough review and evaluation of a standardised, psychometrically valid tool such as the DRDP. The purposes of the working group would be to:

1. Review the DRDP in terms of coverage and developmental alignment with EYLF principles and learning outcomes for children in the year before school, and with expectations of school readiness as children transition to kindergarten. This could form part of a social validity analysis which evaluates criteria such as authenticity, sensitivity, collaboration and utility (see Bagnato et al., 2014).
2. Assess acceptability and relevance of the tool to the everyday work of ECE teachers. Teacher perceptions are essential to successful implementation of a new model for formative assessment, and the selected / adapted tool needs to be seen as relevant to everyday work and as not adding substantial additional load. Ideally, such a tool would reduce load and feed directly into other work that might need to be completed (e.g., end of year summaries; transition to school statement). Evaluation of educator perceptions of the tool should be part of the social validity analysis.
3. Investigate possibilities for aligning the use of the tool within current practices for observing, recording and collecting evidence of children's learning experiences and interests in ECE services. The working group would also need to consider whether the examples within the domains represent readily observable behaviours across different early childhood settings and for diverse groups of children. Providing advice on implementation across diverse populations and services is a key component for effective use of formative assessment.
4. Implement a co-design process with researchers, practitioners, and the tool-developers to generate accurate and useful individualised child reports for teachers to use. The reports should support teachers in making decisions about individual and small group learning needs, and ideally should provide advice on learning experiences that can be undertaken to support children's continued learning and development.
5. Establish psychometrically valid and reliable domain scale-scores for the tool with an Australia sample to test for differential item functioning and/or to establish local standardised developmental norms.
6. Design and conduct implementation research associated with fidelity of tool use, impact on practice in the classroom, impact on outcomes for children, and the professional learning to support tool use. Note that evidence-based decisions about the impact of using a FA tool would need a comparison with a service not using FA (or using a different form of FA) and would need to control for other confounding factors.

PROFESSIONAL LEARNING

There was very little evidence from the Stakeholder Consultations that practitioners had received professional development or training in the use of formative assessment tools or how to interpret and use the information these provide. The Literature Review also highlighted the need and desire amongst practitioners for professional learning and support in the use and interpretation of formative assessment outcomes.

FURTHER RESEARCH AREAS

We believe that pre-service and in-service training should include the need to consider evidence (or lack of) related to the implementation of different formative assessment tools to allow practitioners to think critically and creatively about their choice of tools.

Implementation of ongoing professional learning is needed to support all aspects of formative assessment – appropriate documentation, evaluation of data, and subsequent reflection of individualised teaching, as well as training in the specific implementation of the chosen assessment tool. Professional learning should be in-

depth and active (as opposed passive one-off events) with appropriate recognition given for the engagement in such training.

Specialist professional learning is needed for educational leaders and Centre Directors to understand the expectations for leadership when planning and implementing formative assessment that meets NQS requirements as well as considering the culture of the service and community school culture. Leaders also need opportunities to collaborate with one another across services and communities in an effort to share ideas, strategies, and successes.

The evaluation of professional learning for assessment is also needed, particularly in order to track changes or improvements to teaching strategies.

POLICY ISSUES IN THE IMPLEMENTATION OF FORMATIVE ASSESSMENT

A consistent theme, in both the Stakeholder Consultation interviews and the Literature Review, was educators' concern about the time-consuming nature of recording assessments. Staff who were interviewed mentioned doing this work outside their paid hours. Similarly, in a number of published papers, educators commented that documenting their assessments was a constant burden and took time away from reflection on what they had observed. This defeats the very purpose of assessment, which is to prompt reflection and improve planning and teaching.

On the other hand, the stakeholder consultations and case study visits identified models that were designed to address the challenge of time. In one service, part of the day was spent in child-directed play and activities and part of the day was spent with more teacher-led small group purposeful activities that were designed to better ascertain a child's current and potential level of functioning. Other staff referred to funding they had to employ an additional staff member, which gives teachers more time for one-on-one assessments. Services also work with Early Intervention, speech pathology, occupational therapy and other agencies, which supports their assessment and learning processes.

FURTHER RESEARCH AREAS

We believe that services can be supported to find ways to give ECE practitioners the time to plan diverse activities and compile records on individual children. Increased planning time, both daily and weekly, in combination with the knowledge gained from training on key developmental indicators for activity planning, is key for improving the quality of activities in Early Years environments. Research studies are needed to examine possible options and to evaluate effects on the quality of the planning and practice.

A further area to investigate is the involvement of an extra specialist support worker or liaison officer. The case study services suggest that the presence of specialist support staff can facilitate closer engagement with more children in smaller groupings and allow for more individualised planning. This model is worthy of further research.

REFERENCES

All reviewed work is reported here, even if it is not cited within the main body of this report

- Aguauo, K. B., & Kaminski, R. A. (2012). *Preschool Early Literacy Indicators (PELI): Initial reliability and validity*. Paper presented at the Council for Exceptional Children Annual Convention 2012. Retrieved from https://acadiencelarning.org/papers/PELI_poster_CEC_2012.pdf
- Allal, L. (2002). The assessment of learning dispositions in the classroom. *Assessment in Education*, 9, 55-58
- Australian Children's Education and Care Quality Authority (ACECQA) (2019). *NQF Snapshot Q3 2019*. Retrieved from https://www.acecqa.gov.au/sites/default/files/2019-11/NQFSnapshot_Q32019.pdf
- Australian Children's Education and Care Quality Authority (ACECQA) (2018). *2018 National Quality Standard Assessment and Rating Instrument*. Retrieved from https://www.acecqa.gov.au/sites/default/files/2018-01/NQS_AssessmentAndRatingInstrument.pdf
- Bagnato, S. J., Neisworth, J. T., & Pretti-Frontczak, K. (2014). *LINKing authentic assessment and early childhood intervention: Best measures for best practices*. Paul Brookes Publishing, Baltimore.
- Barghaus, K. M., & Fantuzzo, J. W. (2014). Validation of the preschool Child Observation Record: Does it pass the test for use in Head Start? *Early Education and Development*, 25, 1118-1141. doi:10.1080/10409289.2014.904646
- Blaiklock, K. (2008). A critique of the use of learning stories to assess the learning dispositions of young children. *New Zealand Research in Early Childhood Education*, 11, 77-87.
- Blaiklock, K. (2013). What are children learning in early childhood education in New Zealand? *Australian Journal of Early Childhood*, 38, 51-56.
- Bohlmann, N. L., Downer, J. T., Williford, A. P., Maier, M. F., Booren, L. M., & Howes, C. (2019). Observing children's engagement: Examining factorial validity of the inCLASS across demographic groups. *Journal of Applied Developmental Psychology*, 60, 166-176. doi:10.1016/j.appdev.2018.08.007
- Buldu, M. (2010). Making learning visible in kindergarten classrooms: Pedagogical documentation as a formative assessment technique. *Teaching and Teacher Education*, 26, 1439-1449.
- Burks, C. C. (2018). *Assessment in early childhood intervention: The importance of training*. University of Kentucky.
- Carr, M. (1998b). *Assessing children's learning in early childhood settings: A professional development programme for discussion and reflection*. Wellington, New Zealand Council for Educational Research.
- Carr, M. (2001). *Assessment in early childhood settings: Learning Stories*. London: Paul Chapman.
- Chen, S.-C. S., & Cheng, Y.-P. (2011). Implementing curriculum-based learning portfolio: A case study in Taiwan. *Early Child Development and Care*, 18, 149-164. doi:10.1080/03004430.2011.536637
- Cloney, D., Jackson, J., & Mitchell, D. (2019). *Assessment of children as confident and involved learners in early childhood education and care: Literature review*. VCAA.
- Collado, C. L. (2015). *A case study of formative assessment processes in preschool special education settings*. University of Illinois at Chicago.
- Curenton, S. M., Sims, J., Rochester, S. E., & Gardner, S. L. (2019). The Conversation Compass® Communication Screener-Revised. *Early Childhood Research Quarterly*, 47, 182-193. doi:10.1016/j.ecresq.2018.10.013
- Department of Education, Employment and Workplace Relations (DEEWR) (2009). *Belonging, being and becoming: The Early Years Learning Framework for Australia*. Canberra: DEEWR.
- Department of Education, Employment and Workplace Relations (DEEWR) (2010). *Belonging, being and becoming: Educators Guide to the Early Years Learning Framework for Australia*. Canberra: DEEWR.
- Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*, 25, 1-16. doi:10.1016/j.ecresq.2009.08.004
- DRDP Collaborative Research Group. 2018. Technical Report for the Desired Results Developmental Profile (2015). Report prepared for the California Department of Education.
- Ertle, B., Rosenfeld, D., Presser, A. L., & Goldstein, M. (2016). Preparing preschool teachers to use and benefit from formative assessment: The Birthday Party Assessment professional development system. *ZDM - Mathematics Education*, 48, 977-989. doi:10.1007/s11858-016-0785-9

- Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start Children. *American Educational Research Journal* 48 (3), 763–793 DOI: 10.3102/0002831210385446
- Feldman, E. N. (2010). Benchmarks Curricular Planning and Assessment Framework: Utilizing standards without introducing standardization. *Early Childhood Education Journal*, 38, 233-242. doi:10.1007/s10643-010-0398-9
- Gao, X., & Grisham-Brown, J. (2011). The use of authentic assessment to report accountability data on young children's language, literacy, and pre-math competency. *International Education Studies*, 4, 41-53. doi:10.5539/ies.v4n2p41
- Gettinger, M., & Stoiber, K. C. (2012). Curriculum-based early literacy assessment and differentiated instruction with high-risk preschoolers. *Reading Psychology*, 33, 11-46. doi:10.1080/02702711.2012.630605
- Ginsburg, H., & Pappas, S. (2016). Invitation to the birthday party: Rationale and description. *ZDM - Mathematics Education*, 48, 947-960. doi:10.1007/s11858-016-0818-4
- Gokiert, R., Noble, L., & Littlejohns, L. B. (2013). Directions for professional development: Increasing knowledge of early childhood measurement. *Dialog*, 16, 1-20.
- Goodine, B. R. (2013). *Painting a portrait: Early learning assessment practices*. University of Prince Edward Island.
- Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., Perry, L., & Roberts, M. (2018). *Through Growth to Achievement: Report of the Review to Achieve Educational Excellence in Australian Schools*. Canberra, Australian Government. Retrieved from <https://www.appa.asn.au/wp-content/uploads/2018/04/20180430-Through-Growth-to-Achievement-Text.pdf>
- Greenwood, C. R., Abbott, M., Beecher, C., Atwater, J., & Petersen, S. (2017). Development, validation, and evaluation of Literacy 3D: A package supporting Tier 1 preschool literacy instruction implementation and intervention. *Topics in Early Childhood Special Education*, 37, 29-41. doi:10.1177/0271121416652103
- Hadley, F., Waniganayake, M. & Shepherd, W. (2015) Contemporary practice in professional learning and development of early childhood educators in Australia: Reflections on what works and why, *Professional Development in Education*, 41 (2), 187-202.
- Harrison, L. J., & Wang, A. (2017). *Review of instruments measuring and assessing children's early development*. Unpublished report. Goodstart Early Learning.
- Hatherly, A., & Sands, L. (2002). So what is different about Learning Stories. *The First Years: Nga Tau Tuatahi. New Zealand Journal of Infant and Toddler Education*, 4, 8-12.
- Hojnoski, R., & Floyd, R. G. (2013). *Technical information summary sheet: Individual Growth & Development Indicators of Early Numeracy (1st Edition) – Universal Screening*. Retrieved from: <https://www.myigdis.com/wp-content/uploads/2013/09/IGDIs-EN-1st-Ed-Technical-Info-Summary-Sheet.pdf>
- Hooker, T. (2017). Transforming teachers' formative assessment practices through ePortfolios. *Teaching and Teacher Education*, 67, 440-453. doi:10.1016/j.tate.2017.07.004
- Hooker, T. (2019). Using ePortfolios in early childhood education: Recalling, reconnecting, restarting and learning. *Journal of Early Childhood Research*, 1-16. doi:10.1177/1476718X19875778
- Hostyn, I., Mäkitalo, A. R., Hakari, S., & Vandenbussche, L. (2018). The professional actuation of pedagogical documentation in Belgian and Finnish early childhood education settings. *Early Child Development and Care*, 0(0), 1-14. doi:10.1080/03004430.2018.1475368
- Howard, S. J., Neilsen-Hewett, C., de Rosnay, M., Vasseleu, E., & Melhuish, E. (2019). Evaluating the viability of a structured observational approach to assessing early self-regulation. *Early Childhood Research Quarterly*, 48, 186-197. doi:10.1016/j.ecresq.2019.03.003
- Johnson, L. A. D. (2017). Exploring cloud computing tools to enhance team-based problem solving for challenging behavior. *Topics in Early Childhood Special Education*, 37, 176-188. doi:10.1177/0271121417715318
- Kaminski, R. A., Abbott, M., Bravo Aguayo, K., Latimer, R., & Good, R. H. (2014). The Preschool Early Literacy Indicators: Validity and benchmark goals. *Topics in Early Childhood Special Education*, 34, 71-82. doi:10.1177/0271121414527003
- Kim, D. H., Lambert, R. G., & Burts, D. C. (2013). Evidence of the validity of Teaching Strategies GOLD® assessment tool for English language learners and children with disabilities. *Early Education and Development*, 24, 574-595. doi:10.1080/10409289.2012.701500

- Kim, D. H., Lambert, R. G., Durham, S., & Burts, D. C. (2018). Examining the validity of GOLD® with 4-year-old dual language learners. *Early Education and Development, 29*, 477-493. doi:10.1080/10409289.2018.1460125
- Kim, K. (2016). Teaching to the data collection? (Un)intended consequences of online child assessment system, "Teaching strategies GOLD". *Global Studies of Childhood, 6*, 98-112. doi:10.1177/2043610615627925
- Kim, K. (2018). Early childhood teachers' work and technology in an era of assessment. *European Early Childhood Education Research Journal*. doi:10.1080/1350293X.2018.1533709
- Krause, J. A. (2016). *Exploring early childhood classroom teachers' experiences with administrative support in the implementation of the DRDP as an authentic assessment tool*. University of La Verne.
- Lambert, R., Kim, D. H., & Burts, D. C. (2013). *Evidence for the association between scores from the Teaching Strategies GOLD assessment system and information from direct assessments of child progress*. Retrieved from <https://ceme.uncc.edu/sites/ceme.uncc.edu/files/media/pdfs/Concurrent%20Validity%20final.pdf>
- Lambert, R. G., Kim, D. H., & Burts, D. C. (2014). Using teacher ratings to track the growth and development of young children using the Teaching Strategies GOLD assessment system. *Journal of Psychoeducational Assessment, 32*, 27-39. doi:10.1177/0734282913485214
- Lambert, R. G., Kim, D.-H., & Burts, D. C. (2015). The measurement properties of the Teaching Strategies GOLD assessment system. *Early Childhood Research Quarterly, 33*, 49-63. doi:10.1016/j.ecresq.2015.05.004
- Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*, 448-465. doi:10.1037/a0013842
- Landry, S. H., Swank, P. R., Anthony, J. L., & Assel, M. A. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Reading and Writing: An Interdisciplinary Journal, 24*(8), 971-1010.
- Landry, S., Assel, M. A., Williams, J. M., Zucker, T., Swank, P. R., & Gunnewig, S. (2014). *CIRCLE (formerly C-PALLS+STEM) Technical manual*. Retrieved from <https://hisdearlychildhood.files.wordpress.com/2013/08/circle-technicalmanual-2.pdf>
- Laracy, S. D., Hojniski, R. L., & Dever, B. V. (2016). Assessing the classification accuracy of early numeracy curriculum-based measures using receiver operating characteristic curve analysis. *Assessment for Effective Intervention, 41*, 172-183. doi:10.1177/1534508415621542
- Lee, Y. S. (2016). Psychometric analyses of the Birthday Party. *ZDM - Mathematics Education, 48*, 961-975. doi:10.1007/s11858-016-0801-0
- Little, M., Cohen-Vogel, L., Sadler, J., & Merrill, B. (2019). Data-driven decision making in early education: Evidence from North Carolina's Pre-K program. *Education Policy Analysis Archives, 27*, 1-23. doi:10.14507/epaa.27.4198
- Loggenberg, E. M. (2011). *Assessment in early childhood education in New Zealand*. Massey University.
- Marbina, L., Mashford-Scott, A., Church, A., & Taylor, C. (2015). *Assessment of wellbeing in early childhood education and care: Literature review*. VCAA.
- McConnell, S., Bradfield, T. A., Wackerle-Hollman, A. K., & Rodriguez, M. C. (2012). *Technical information summary sheet: Individual Growth & Development Indicators of Early Literacy (1st Edition)*. Retrieved from <https://www.myigdis.com/wp-content/uploads/2013/01/IGDIs-EL-2nd-Ed-Technical-Info-Summary.pdf>
- McLean, L. (2018). *Co-construction of pedagogical documentation by children and educators in early learning environments*. (Doctor of Philosophy). University of Toronto
- Missall, K. N., McConnell, S. R., & Cadigan, K. (2006). Early literacy development: Skill growth and relations between classroom variables for preschool children. *Journal of Early Intervention, 29*(1), 1-21. doi:10.1177/105381510602900101
- Moiduddin, E., Kamler, C., Malone, L., & Gonzalez, K. (2014). *Milpitas Early Learning Transitions model: Using assessment data to inform teacher practice*.
- Nguyen, T., Reich, S. M., Jenkins, J. M., & Abedi, J. (2019). Psychometric validation and reorganization of the Desired Results Developmental Profile. *Journal of Psychoeducational Assessment, 1-20*. doi:10.1177/0734282919859804
- Niles, A. J. (2016). *Complexities of assessment: Striving to get it 'right'*. (Masters of Education). University of Canterbury.
- Nolan, A., Morrissey, A., and Dumenden, I. (2013). Expectations of mentoring in a time of change: views of new and professionally isolated early childhood teachers in Victoria, Australia. *Early Years, 33* (2), 161-171.

- NSW Dept of Education (2018). *Transition to School Statement (T2S Statement)*. Retrieved from <https://education.nsw.gov.au/early-childhood-education/working-in-early-childhood-education/transition-to-school/T2S-Statement-interactive-form-2018.pdf>
- NSW Education Standards Authority (NESAs) (2018). *Australian Professional Standards for Teachers*. Retrieved from www.nesa.nsw.edu.au
- Nyland, B., & Alfayez, S. (2012). Learning stories – crossing borders: Introducing qualitative early childhood observation techniques to early childhood practitioners in Saudi Arabia. *International Journal of Early Years Education*, 20, 392-404. doi:10.1080/09669760.2012.743097
- Patterson, C., & Fleet, A. (2011). *Planning in the context of the EYLF: powerful, practical and pedagogically sound*. (Research in practice series; Vol. 18, No. 2). Deakin West, A.C.T.: Early Childhood Australia.
- Piasta, S. B., Phillips, B. M., Williams, J. M., Bowles, R. P., & Anthony, J. L. (2016). Measuring young children's alphabet knowledge: Development and validation of brief letter-sound knowledge assessments. *The Elementary School Journal*, 116, 523-548.
- Piper, A. M., D'Angelo, S., & Hollan, J. D. (2013). *Going digital: Understanding paper and photo documentation practices in early childhood education*. Paper presented at the Computer Supported Young People.
- Purpura, D. J., & Lonigan, C. J. (2015). Early numeracy assessment: The development of the preschool Early Numeracy Scales. *Early Education and Development*, 26, 286-313. doi:10.1080/10409289.2015.991084
- Redding, C., Cannata, M., & Taylor Haynes, K. (2017). With scale in mind: A continuous improvement model for implementation. *Peabody Journal of Education*, 92, 589-608.
- Riley-Ayers, S., Frede, E. C., & Jung, K. (2010). *Early Learning Scale: Technical report*. National Institute for Early Education Research. Retrieved from http://nieer.org/wp-content/uploads/2012/03/Early_learning_scale_tech_report_Sept_2010.pdf
- Roseth, C. J., Missall, K. N., & McConnell, S. R. (2012). Early Literacy Individual Growth and Development Indicators (EL-IGDIs): Growth trajectories using a large, internet-based sample. *Journal of School Psychology*, 50, 483-501. doi:10.1016/j.jsp.2012.03.001
- Rowe, D. W., & Wilson, S. J. (2015). The development of a descriptive measure of early childhood writing: Results from the Write Start! writing assessment. *Journal of Literacy Research*, 47, 245-292. doi:10.1177/1086296X15619723
- Sheridan, S., Williams, P., & Sandberg, A. (2013). Systematic quality work in preschool. *International Journal of Early Childhood*, 45, 123-150. doi:10.1007/s13158-012-0076-8
- Singer, C. R. (2009). *The use of the Child Observation Record as a tool for predicting early literacy skills in preschool children*. Rochester Institute of Technology.
- Siraj, I., Melhuish, E., Howard, S., Neilsen-Hewett, C., Kingston, D., de Rosnay, M., Duursma, E., Feng, X. & Luu, B. (2018). *Fostering Effective Early Learning (FEEL) Study: Final Report*. Sydney, Australia: NSW Department of Education. <https://education.nsw.gov.au/early-childhood-education/whats-happening-in-the-early-childhood-education-sector/data-andresearch/feel-study-2018>
- Sisson, J. & Whittington, V. (2018). Pedagogical documentation: A South Australian Perspective: A resource for early education and care educator teams. <https://gowriesa.org.au/docs/Pedagogical-Documentation-08-02-2018.pdf>
- Solari, E. J., Zucker, T. A., Landry, S. H., & Williams, J. M. (2016). Relative effects of a comprehensive versus reduced training for Head Start teachers who serve Spanish-speaking English learners. *Early Education & Development*, 27, 1060-1076. doi:10.1080/10409289.2016.1158610
- Spencer, T. D., Goldstein, H., Kelley, E. S., Sherman, A., & McCune, L. (2017). A curriculum-based measure of language comprehension for preschoolers: Reliability and validity of the Assessment of Story Comprehension. *Assessment for Effective Intervention*, 42, 209-223. doi:10.1177/1534508417694121
- Susman-Stillman, A., Bailey, A. E., & Webb, C. (2014). *The state of early childhood assessment: Practices and professional development in Minnesota*. Retrieved from <http://ceed.umn.edu/wp-content/uploads/2017/04/The-state-of-early-childhood-assessment-practices-and-professional-development-in-minnesota.pdf>
- Taylor, B. S. (2018). *Scoring reliability by early childhood educators on a curriculum based assessment*. University of Kentucky.
- Teaching Strategies GOLD assessment system: Concurrent validity*. (2013). Retrieved from <https://teachingstrategies.com/our-approach/research/>
- Verdon, S., et al. (2018). *Assessment of children as effective communicators in early childhood education and care: Literature review*. Melbourne: VCAA

- Wakabayashi, T., Claxton, J., & Smith, E. V. (2019). Validation of a revised observation-based assessment tool for children birth through kindergarten: The COR Advantage. *Journal of Psychoeducational Assessment, 37*, 69-90. doi:10.1177/0734282917732491
- Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education – Or whose score is it anyway? *Early Childhood Research Quarterly, 27*, 46-54. doi:10.1016/j.ecresq.2011.06.003
- Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., . . . Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4*, 63-85. doi:10.1007/s11409-008-9033-1
- Williford, A. P., Bulotsky-Shearer, R., Bichay, K., Reilly, S., & Downer, J. T. (2018). Adapting assessments of child engagement to develop an early childhood consultation model. *Journal of Applied Developmental Psychology, 56*, 67-78. doi:10.1016/j.appdev.2018.01.005
- Winchell, B. (2011). *A critical examination of the technical adequacy of a curriculum-based assessment using Rasch analysis*. Kent State University College.
- Yunus, F. (2014). Cross-cultural adaptation of developmental criteria for young children: A preliminary psychometric study. *International Education Studies, 7*, 130-145. doi:10.5539/ies.v7n13p130
- Zhang, Q. (2017). Do learning stories tell the whole story of children's learning? A phenomenographic enquiry. *Early Years, 37*, 255-267. doi:10.1080/09575146.2016.1151403

APPENDIX A: LITERATURE SEARCH KEYWORDS

A. **Formative Assessment** keywords:

"formative assessment" OR "formative learning*" OR "formative teach*" OR "assessment for learning" OR "diagnostic assessment*" OR "formative evaluation*" OR "mastery learning" OR "progress monitoring" OR "learning progress assessment" OR "teacher feedback" OR "learner feedback" OR "informed pedagogy" OR "informed practice" OR "ongoing cycle" OR "critical reflection"*

B. **Assessment Tool** keywords:

"assessment tool" OR "assessment instrument*" OR "assessment system*" OR "monitoring tool*" OR "embedded measure*" OR "learning progression" OR "classroom questioning" OR "curriculum-based assessment*" OR "curriculum-based measurement" OR "curriculum-embedded measure" OR "curriculum-embedded measurement"*

C. **Early Childhood** keywords:

"early childhood" OR "early years" OR "preschool" OR "pre-school" OR kindergarten OR childcare OR "child care" OR "long day care" OR "family day care" OR "family child care" OR "day care" OR "early education" OR "occasional care" OR playgroup OR creche* OR "year before school" OR "pre-K" OR "pre-kindergarten"*

D. **Professional Development** keywords:

"professional learning" OR "professional development" OR "communit of practice" OR "professional learning communit*" OR "professional network*" OR "teacher network*" OR "online communit*" OR train* OR "professional support" OR "pedagogical support"*

E. **Supports/Barriers** keywords:

"role of teacher" OR leader* OR principal* OR "educational leader*" OR "pedagogical leader*" OR "cent* director*" OR "room leader*" OR admin* OR "role of admin*" OR "role of headm*" OR "support resource*" OR "continuous improvement" OR "quality improvement"*

F. **Australia** keywords:

"Australia" OR "NSW" OR "New South Wales" OR "Queensland" OR "QLD" OR "Victoria" OR "VIC" OR "South Australia" OR "SA" OR "Western Australia" OR "WA" OR "Tasmania" OR "TAS" OR "Northern Territory" OR "NT" OR "Indigenous" OR "Aboriginal" OR "Torres Strait"*

G. **Tool Name** keywords: In addition, specific tool searches were also conducted using Google Scholar. As Google Scholar does not support Boolean logic to the same extent as academic databases, we generally searched for the tool name or for papers citing the paper where the tool was initially published.

APPENDIX B: LIST OF FORMATIVE ASSESSMENT TOOLS

Domain-general tools						
Tool name	Source	Match to EYLF outcomes				
		Outcome 1: Children have a strong sense of identity	Outcome 2: Children are connected with and contribute to their world	Outcome 3: Children have a strong sense of wellbeing	Outcome 4: Children are confident and involved learners	Outcome 5: Children are effective communicators
Assessment, Evaluation, and Programming System 2nd and 3rd edition (AEPS-2/3)	https://brookespublishing.com/product/aeps/	✓	✗	✓	✓	✓
Benchmarks curricular planning and assessment framework (BCPAF)	Feldman (2010)	✓	✓	✓	✓	✓
CIRCLE, formerly C-PALLS + STEM (The Phonological Awareness Language and Literacy System + Science, Technology, Engineering & Math)	https://cliengage.org/public/tools/assessment/circle-progress-monitoring/	✗	✗	✓	✓	✓
Checklist of Independent Learning Development 3–5 (CHILD 3–5)	Whitebread et al., 2009	✗	✗	✓	✓	✓
Child Observation Record (COR)	https://highscope.org/cor-advantage/	✓	✓	✓	✓	✓
Desired Results Developmental Profile (DRDP)	https://www.desiredresults.us/drdp-forms	✓	✓	✓	✓	✓
Early Learning Scale (ELS)	https://www.myelsoonline.com/www/index.php	✗	✗	✓	✗	✓
Individualized Classroom Assessment Scoring System (InCLASS)	http://www.inclassobservation.com/	✗	✗	✓	✓	✓
Learning Stories	Carr (1998, 2001); Updated publication Carr & Lee (2012).	✓	✓	✓	✓	✓
Profile of Preschool Learning & Developmental Readiness (ProLADR)	https://www.myigdis.com/preschool-assessments/socia	✗	✗	✓	✓	✓

	l-emotional-assessments/					
Self-designed (Behavior Incident Reporting System via Google Sheets)	Johnson (2017)	✗	✗	?	?	?
Teaching Strategies GOLD	https://teachingstrategies.com/solutions/assess/gold	✗	✗	✓	✓	✓

Domain-specific tools						
Tool name	Source	Match to EYLF outcomes				
		Outcome 1: Children have a strong sense of identity	Outcome 2: Children are connected with and contribute to their world	Outcome 3: Children have a strong sense of wellbeing	Outcome 4: Children are confident and involved learners	Outcome 5: Children are effective communicators
Assessment of Story Comprehension	Spencer et al., (2017)	✗	✗	✗	✗	✓
The Birthday Party	Ginsburg & Pappas (2016)	✗	✗	✗	✓	✗
Conversation Compass Communication Screener - Revised (CCCS-R)	Curenton et al., 2019	✗	✗	✗	✗	✓
Early Literacy - Individual Growth and Development Indicators (EL-IGDIs)	https://www.myigdis.com/preschool-assessments/early-literacy-assessments/#1460350206401-de3ecdce-928f	✗	✗	✗	✗	✓
Early Numeracy Scales	Purpura & Lonigan (2015)	✗	✗	✗	?	✗
Letter-Sound Short Forms	Piasta et al., (2016)	✗	✗	✗	✗	✓
Numeracy - Individual Growth and Development Indicators (N-IDGIs)	https://www.myigdis.com/preschool-assessments/early-numeracy-assessments/	✗	✗	✗	✓	✗
Preschool Early Literacy Indicators (PELI)	https://acadiancelearning.org/peli.html	✗	✗	✗	✗	✓
Preschool Situational Self-Regulation Toolkit (PRERSIST)	http://www.prsist.com.au/	✗	✓	✗	✓	✗
Write Start! Writing Assessment	Rowe & Wilson 2015	✗	✗	✗	✗	?

APPENDIX C: DETAILED DESCRIPTIONS OF FORMATIVE ASSESSMENT TOOLS AND SUMMARIES OF RESEARCH EVIDENCE

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Assessment, Evaluation, and Programming System 2nd 3rd ed. AEPS-2 and -3</p> <p>AEPS is designed for determining a child's current level of functioning based on their performance of everyday activities, developing developmentally appropriate goals for a child, planning intervention, and evaluating performance over time.</p> <p>Latest version: 2018</p> <p>Source: https://brookespublishing.com/product/aeps/</p> <p>Age range: 0:0–6:0 years</p> <p>Cost: \$179/set</p> <p>Number of items: 217 items: 15 fine motor, 17 gross motor, 35 adaptive, 54 cognitive, 49 social communication, 47 social.</p> <p>Time: Not specified, "during regular classroom activities".</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Online data entry system available (AEPSi).</p> <p>Standardised scoring: Yes</p> <p>Training: No pre-requisite online training offered; provide technical assistance including development of protocols, teacher training, onsite support. Offer inter-rater reliability training for a fee.</p> <p>Who administers the tool and receives feedback: Educators; family can report home observations.</p> <p>Details included in feedback: Two scores for cognitive and social-communication domains, indicating how consistently the rubrics are met; a section for individualised comments.</p> <p>Suitable for diverse populations: AEPS-2 is not suited to measure small progress in</p>	<p>1. Sense of identity Social subtest includes interaction with others, participation, interaction with environment, knowledge of self and others.</p> <p>3. Sense of well-being Fine and gross motor movement.</p> <p>4. Confident and involved learners Cognitive domain.</p> <p>5. Effective communication Cognitive, social, and social-communication domains.</p>	<p>Inter-rater reliability: With 1.5 days of training, 6 graduate students achieved an excellent inter-rater reliability of >80% when using AEPS with Spanish-speaking children living in a Guatemalan group children's home. Assessment was conducted in English via a translator (Taylor, 2018).</p> <p>Level II of AEPS (3–6 years old) demonstrated good reliability as assessed by Rasch analysis, with all item separation reliability values reaching 1.00 (Winchell, 2011). This indicates that item difficulty hierarchy is working as expected.</p>	<p>Construct validity: Gao et al., (2011) report medium-to-high concurrent validity between AEPS-2 Social-communication subscale and BDI-2 Communication subscale, as well as AEPS-2 cognitive and BDI-2 cognitive subscales.</p>	<p>Confirmatory factor analysis identified six factors corresponding to intended domains, accounting for 64.3% of total variance explained (Winchell 2011).</p>	<p>Social validity: Teachers surveyed by Gao et al., (2011) commented that AEPS is easier for children than standardised tests. All teachers reported using AEPS results to inform instruction, trusted test results, and believed that the assessed skills are linked to both daily life and the curriculum.</p> <p>Developmental and subgroup validity: Rasch analysis on a large dataset suggested that two AEPS items are more difficult for males than females and five items are more difficult for females than males (Winchell, 2011). 45 items showed developmental bias, such that 14 are more difficult for children with developmental delays than typically developing children. 18 items are more</p>	NA

Tool			Psychometrics			Evidence	
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
	younger children, especially those at-risk or experiencing developmental delays (Winchell, 2011). AEPS-3 includes more examples to address cultural diversity and has been used to report on children with a wide age range and disability status (Taylor, 2018). It had weaker inter-rater reliability in Malaysian context (Yanus, 2014).					difficult for at-risk children than for typically developing children, and 13 are more difficult for children with developmental delays than for children at-risk. Gao et al., (2011) only used cognitive and social communication scales, Winchell (2011) used all subscales.	
Assessment of Story Comprehension	Intended cycle frequency: For progress-monitoring, 'frequent' administration of 1 parallel item. For decision-making, 3 items at 'specific times of the year'. Online or IT-based mode: No Standardised scoring: Not by default (mean differences between items). Training: RAs received 2h training for administration and scoring Who administers the tool and receives feedback: Validated with RA, intended to be accessible for educators.	5. Effective communication Language comprehension	Inter-rater reliability: Kappa coefficients for each item ranged from .60 to .94 suggesting moderate to high scoring reliability (Spencer et al., 2017). Intra-rater reliability: 86% of variation between item scores is due to child-level differences, leaving only 14% for difference between raters, which is considered good. Internal reliability: Mean consistency across all nine forms was reasonably high (Cronbach's alpha = .83). Mean Cronbach's alpha for 8 items on each form was .81,	Construct validity: ASC had moderate to large correlations with a standardised test of oral language abilities (CELF-P) and a test of narrative language (TNR) (Spencer et al., 2017).	NA	Data from a pilot study suggested that the assessment was sufficiently sensitive to detect change following intervention (Spencer et al., 2017).	NA
ASC intends to identify children who could benefit from supplemental language intervention and to monitor language growth of children who participate in language intervention. Latest version: 2017 Source: Spencer et al., (2017) Age range: 3:0–5:0 years Cost: Not a commercial tool Number of items: 9 parallel items Time: 5 minutes							

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
	<p>Details included in feedback: Rubric scores</p> <p>Suitable for diverse populations: Initial validation with children from Head Start (low income background).</p>		<p>ranging from .71 to .89.</p> <p>Parallel form reliability: Correlations between different ASC forms ranged from medium to high (.65–.83), but significant t-tests between some forms suggest they are not equivalent.</p>				
<p>Benchmarks curricular planning and assessment framework (BCPAF)</p> <p>Provides criteria for selecting benchmarks during planning and translating teachers' observations into early learning standards as part of assessment.</p> <p>The educator develops/chooses a list of benchmarks that they think will be met during a particular activity. Then the activity is observed and analysed within this framework. The benchmarks identified for individual children are summarised to estimate which benchmarks are being met by how many children.</p> <p>Latest version: 2010</p> <p>Source:</p>	<p>Intended cycle frequency: Not specified</p> <p>Online or IT-based mode: Can videotape observations for better reliability.</p> <p>Standardised scoring: No</p> <p>Training: Research assistants were trained to use the tool, no commercial variant available.</p> <p>Who administers the tool and receives feedback: RAs, potentially suitable for educators.</p> <p>Details included in feedback: No feedback other than the list of benchmarks being met by the children. No scaffolding for changing instruction is provided.</p> <p>Suitable for diverse populations: NA</p>	<p>Aligned with Washington State learning standards, the approach can be adapted to EYLF outcomes.</p>	<p>Inter-rater reliability: Feldman (2010) calculated reliability rates using one live observation session (78% agreement, Kappa = .52, $r = .62$) and eight videos of children playing in a variety of settings (83% mean agreement, range: 76–88%; Kappa = .48, range: .34–.68; mean $r = .52$, range: .22 –.77). These statistics suggest moderate inter-rater reliability.</p> <p>Intra-rater reliability: Reliability rates were calculated based on coders' rescoring of six videos of the original eight videos (average 83% agreement, range: 76–88%; mean Kappa = .48, range: .34–.68; mean $r = .52$, range:</p>	NA	NA	<p>Feldman (2010) describe the framework as an improvement on authentic assessment.</p>	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Feldman (2010)</p> <p>Age range: 0:10–8:0 years (validated to “Kindergarten” age)</p> <p>Cost: Unclear</p> <p>Number of items: NA</p> <p>Time: Not specified – looks time-consuming but can be done for the entire class at once if videotaping, no effort on the part of the children required.</p>			.22–.77). These suggest moderate intra-rater reliability.				
<p>The Birthday Party</p> <p>The BP was designed to provide formative assessment to guide instruction, evaluate the implementation of a variety of curricula, and can be used for screening children at risk of mathematical difficulties.</p> <p>Results uploaded to a centralised, secure server can be used to provide accountability by comparing classrooms with local sites as well as larger units like districts, states, and indeed the country as a whole. Data regarding these comparisons can be exported to statistical databases for appropriate analyses.</p> <p>Latest version: 2016</p> <p>Source:</p>	<p>Intended cycle frequency: Not specified.</p> <p>Online or IT-based mode: Anything with a touch screen.</p> <p>Standardised scoring: Yes</p> <p>Training: Created workshops and a website designed to enable a wide range of early childhood professionals, including assessment specialists, teachers, and education directors, to understand the BP, to enrich their views of children’s thinking and learning, to administer it comfortably, and to interpret the results and their implications for instruction.</p>	<p>4. Confident and involved learners</p> <p>Number and operations (counting, producing sets of objects, adding and subtracting, comparing numbers, using one-to-one correspondence, and understanding the principles of cardinality and identity);</p> <p>Shape (recognizing shapes and identifying their attributes);</p> <p>Space (understanding words that describe relative location, such as “on” and “left,” and proximity, such as “closest” and “between”;</p> <p>constructing formations of chips on grids, such as vertical,</p>	<p>Internal reliability and test-retest reliability: Cronbach’s alpha coefficients were satisfactory, ranging from .70 to .94 for all measures across all age groups. Test–retest reliability coefficients were highest for the Number and Operation measure across all age groups, with the largest value of .82 for age 4. Generally, the lowest test–retest reliability was observed for the Pattern and Space measures (Lee, 2016).</p> <p>Inter-rater reliability: For the inter-rater reliability measured by shadow scoring, all coefficients were high (>.90) except the</p>	<p>Convergent validity: Positive correlations with a standardised achievement task indicated adequate concurrent and predictive validity (Lee, 2016).</p>	<p>Confirmatory factor analysis confirmed acceptable construct validity for all age groups on an underlying construct of ‘early mathematics proficiency’ (Lee, 2016).</p>	<p>Developmental and subgroup validity: DIF (differential item functioning) analysis suggests that items are functioning the same way regardless of gender, SES, or language spoken at home. Findings from CTT (classical test theory) and IRT (item response theory) item analyses indicate that item difficulty displayed optimal variation and increased both within and across age groups on common items. The progression in difficulty lends empirical support to the theoretically developed items (Lee, 2016).</p>	<p>Erte et al., 2016 (T) – an online survey of 11 early childhood pre-service students. Participants reviewed and evaluated the BP workshops website that aimed to provide training for educators using BP. All respondents identified value to the website with regard to conducting assessments, understanding the purpose of assessments, or what they could learn about mathematics, children’s mathematics learning, or how to teach mathematics. At the same time, the website was not entirely successful in helping respondents use the website ideas to shape specific teaching practices.</p> <p>The PD system is comprised of workshops and an educative website that utilises the power and attraction of videos that illuminate children’s mathematical thinking as they respond to assessment tasks administered via a clinical interview. The PD is intended to help teachers understand what the BP can measure, how they can conduct their</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Ginsburg & Pappas (2016)</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: Varies depending on age. Maximum 36 items across 11 tasks.</p> <p>Time: 20 mins but varies by child age.</p>	<p>Who administers the tool and receives feedback: Educators.</p> <p>Details included in feedback: Minimally adaptive: the tool will revert back to previous instructions if the child is not responding or responds incorrectly. Report provides item-level data for each student, as well as strategy and math language use. Total score is based on accuracy of responses within a domain and across domains.</p> <p>Suitable for diverse populations: The BP was translated into Spanish. The game-like activities, drawing on familiar and interesting events like birthday parties, have been shown to work with many ethnic groups (Ginsburg, Choi, Lopez, Netley, & Chao-Yuan 1997) and should be appropriate for children with learning or emotional difficulties.</p>	<p>horizontal, and diagonal lines); Pattern (matching, copying, and extending patterns).</p>	<p>Pattern and Space measures for the age 3 group (.71 for Pattern and .81 for Space) (Lee, 2016).</p>				<p>own assessments, and how this information can guide teaching.</p>
<p>CIRCLE (formerly C-PALLS+STEM)</p> <p>An online platform for direct assessment of literacy skills, mathematics, science and social studies, and for</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Yes</p> <p>Standardised scoring: Yes</p>	<p>2. Connected with and contribute to the world</p> <p>Mathematics, science, social studies assessments.</p>	<p>Internal reliability: Cronbach's alpha is not appropriate for timed tests, so is only reported for phonological awareness (>.90 for all 3 age groups).</p>	<p>Construct validity: Landry et al., (2014) report that across all age groups, vocabulary subscale had the highest correlation with EOWPVT (.45-.59) and letter subscale has the</p>	<p>Structural validity: CFA on the phonological awareness scale did not support a unidimensional model, but further analysis suggested</p>	<p>Developmental validity: Growth models confirmed that CIRCLE language and literacy subtests are sensitive to child maturation and there were no major ceiling</p>	<p>Landry et al., 2009 (T/S) – 262 preschool educators in the US were trained to administer the tool either online or manually. Children whose teachers received training performed better than controls on all measures, but there was no difference between groups whose teachers received progress monitoring or</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>observation-based assessments of social, emotional, and physical development, as well as early writing, speech production and sentence skills, motivation to read and approaches to learning.</p> <p>Latest version: 2014</p> <p>Source: https://cliengage.org/public/tools/assessment/circle-progress-monitoring/</p> <p>Age range: 3:6–5:0 years</p> <p>Cost: Request quote</p> <p>Number of items: Multiple direct and observational assessments across many domains</p> <p>Time: All assessment tasks take <10 minutes to complete, most take <5 minutes.</p>	<p>Training: Available through the provider.</p> <p>Who administers the tool and receives feedback: Educator, reports for parents available.</p> <p>Details included in feedback: The provider has a collection of instructional activities that target specific skills. Unclear if the suggested activities are specific to a child's score.</p> <p>Suitable for diverse populations: Not evaluated.</p>	<p>3. Sense of well-being Social and emotional development, physical health and development.</p> <p>4. Confident and involved learners Mathematics, motivation to read, and approaches to learning.</p> <p>5. Effective communication Rapid letter naming, letter-sound correspondence, phonological awareness, rapid vocabulary, book and print awareness, story retell and comprehension. Early writing, speech production and sentence skills.</p>	<p>Intraclass correlation coefficients were good across all age groups: .66 for vocabulary, .74 for letters, and .66 for phonological awareness (Landry et al., 2014).</p> <p>Inter-rater reliability: Item-level inter-rater agreement for the science subtests reached 99%. Internal consistency was .81, and test-retest reliability was .82.</p> <p>Test-retest reliability across all age groups was .59-.68 for vocabulary, .66-.83 for letters, and .58-.75 for phonological awareness.</p>	<p>highest correlation with Print Knowledge (from Pre-CTOPP, .76-.79), as expected. Phonological awareness had the strongest (.37-.39) correlation with DSC as compared to Vocabulary and Letter subscales, and it had higher correlations with EOWPVT (.46-.47), and Print Knowledge (.50-.61).</p> <p>Mathematics subtest strongly correlated with the Child Math Assessment ($r = .77$) and moderately correlated with Woodcock-Johnson III Applied Problems subtest (.55). Science subtest strongly correlated with the Preschool Science Assessment (.81).</p> <p>Social and Emotional Scale Checklist had weaker validity, as its subscales had small to medium correlations (.25 - .61) with other established measures (SCBE, Externalizing problems, General Adaptation, Children's Behavior Questionnaire).</p>	<p>that this may be due to the different response format (close-ended vs open-ended) generating different response patterns. Accounting for this is an improved model fit (CFI=.984; TLI=.996; RMSEA=.017).</p> <p>Mathematics and science subtests both fit a unidimensional model, although some of the items were excluded from the science subtest as they were found to have poor discrimination (Landry et al., 2014).</p>	<p>or floor effects (Landry et al., 2014).</p> <p>Psychometric analyses specifically excluded children younger than 3.5yo, although this is not mentioned on the website ("best fit for 3- and 4-year olds").</p> <p>Subgroup validity: CIRCLE was not evaluated on children with disabilities and is not designed to make diagnostic decisions (Landry et al., 2014).</p>	<p>mentoring type of PD. Both intervention groups (online vs paper administration) showed a greater increase in frequency and quality of specific teaching behaviours, and the online administration group improved more than the group administering CIRCLE on paper.</p> <p>Landry et al., 2011 (T/S) – 209 educators received PD focused on classroom management, best practices and responsive teaching (the best form of PD from Landry et al., (2009)). Authors do not specify whether tool-specific training was provided. Children whose teachers received 2 years of PD (vs 1.5 years) had greater vocabulary ($d = .16$), greater gains in complex language development, and print-knowledge (but P-K depended on pre-test scores ($d = .34$). Teachers showed significant gains in observed teacher behaviours ($d = .84$), except for written expression. Receiving 2 vs 1.5 years of PD did not result in significant differences in gains.</p> <p>Solari et al., 2016 (T/S) – 46 educators received either progress-monitoring or progress-monitoring and PD in using data to inform instruction and responsive teaching. Both groups received basic training in using CIRCLE. Compared to the progress-monitoring only group, children in the PD group had significantly improved scores on Spanish letter names ($d = .30$), Spanish PA blending task from Pre-CTOPP ($d = .41$), and English PLS-4, but not on any other subscales. For teachers, there were no significant changes from pre-test to post-test on any subscales of the teacher behaviour rating scale.</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
							Greenwood et al., 2017 (T/S) – 20 educators received introductory PD and 20 hours of coaching spread over the course of the year, focusing on adjusting instruction to CIRCLE and PELI scores. Teachers in the intervention group showed a 16% increase in literacy-related instruction, significantly more than the control group ($d = 1.28$). The instructional quality mean score improved over time within the intervention group, increasing from 66% (SD = 12) in fall and increasing to 86% (SD = 10) in spring, $t(9) = 5.478$, $p < .001$. This improvement covaried with the uptake in evidence-based literacy practices in this group. Throughout the intervention, children showed an upward trend in exhibited literacy skills, gaining 11% on average. Children also showed a significant increase in PELI Composite scores. Children with individual education plans had a higher growth rate than those without IEPs, ($p = .003$), indicating that they were closing the gap in early literacy skills.
<p>Checklist of Independent Learning Development 3–5 (CHILD 3–5)</p> <p>A practical assessment tool for teachers to evaluate children's metacognition and self-regulation.</p> <p>Latest version: 2009</p> <p>Source: Whitebread et al., (2009)</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: No</p> <p>Standardised scoring: No</p> <p>Training: No</p> <p>Who administers the tool and receives feedback: Educators</p> <p>Details included in feedback: A pattern of responses from "Always" to "Never" across items.</p>	<p>3. Sense of well-being Emotional and motivational regulation</p> <p>4. Confident and involved learners Metacognitive regulation</p> <p>5. Effective communication Metacognitive knowledge</p>	<p>Inter-rater reliability: Average inter-rater reliability across four domains was excellent (86%), ranging from 75% on motivational subscale to 91% of social subscale. Untrained raters (classroom nurses) chose an identical response 56% of the time, and only varied by one category 96% of the time (e.g. 'always' or 'usually')</p>	NA	NA	NA	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Age range: 3:2–5:6 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: 22</p> <p>Time: Unspecified</p>	<p>Suitable for diverse populations: Initial validation study used a representative sample of types of pre-school provision and socio-economic catchment area in the Cambridgeshire region.</p>		<p>(Whitebread et al., 2009).</p> <p>Internal reliability: The scale shows high internal consistency, with Cronbach's alpha = .97 for the 22 items.</p>				
<p>Conversation Compass Communication Screener - Revised (CCCS-R)</p> <p>CCCS-R is intended to assess pre-schoolers' discourse skills in the classroom and can help teachers to modify their language practices to meet the needs of their students. It can also be used for research purposes.</p> <p>Latest version: 2019</p> <p>Source: Curenton et al. (2019)</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: 37</p> <p>Time: Unspecified</p>	<p>Intended cycle frequency: Unspecified</p> <p>Online or IT-based mode: No</p> <p>Standardised scoring: Yes</p> <p>Training: No</p> <p>Who administers the tool and receives feedback: Research assistants, educators</p> <p>Details included in feedback: A score for each domain</p> <p>Suitable for diverse populations: The average classroom had 38% Latino students and 24% Black/African American students. Teachers identified 28% (n = 71) of the children as dual language learners.</p>	<p>5. Effective communication</p> <p>Higher order thinking (pre-academic language), negative communication behaviours, social communication, narrative talk and vocabulary knowledge.</p>	<p>Inter-rater reliability: Intraclass correlation coefficients ranged from 0.14 on Descriptive Pragmatics Profile scores to 0.19 on Pre-Literacy Rating Scale scores, suggesting that the majority of variance in outcomes was attributable to differences between classrooms) and therefore that there were only small differences between teachers' scoring of the CCCS-R (Curenton et al., 2019).</p>	<p>Convergent validity: All four CCCS-R domains had significant medium relationships with CELF-P Descriptive Pragmatics Profile, Cohen's d ranging from .43 to .69.</p> <p>Predictive validity: CCCS-R domain scores were significant predictors of Pre-Literacy Rating Scale scores, with small to medium standardised coefficients ranging from .21 to .70 (Curenton et al., 2019).</p>	<p>CFA suggested that the original CCCS did not adequately distribute all items between domains. These items were identified through EFA and the new CCCS-R scale conformed to the theorised 4-factor model. However, fit statistics were still only moderately adequate (.084 RMSEA, CFI = .860 and TLI = .847) (Curenton et al., 2019).</p>	<p>Developmental validity: As suggested in Gardner and Curenton's prior work, the grammar scale was eliminated from the CCCS-R in order to provide a more developmentally appropriate assessment of preschool-aged children's grammar skills, especially those children that are dual language learners or who speak vernaculars such as African American English.</p>	NA
<p>Child Observation Record (COR)</p> <p>Teachers or caregivers spend a few minutes each</p>	<p>Intended cycle frequency: The observational time period is determined by program administrators</p>	<p>1. Sense of identity</p> <p><i>Social and emotional development:</i> emotions, building relationships with</p>	<p>Internal reliability: Content experts supported the usability of COR for teachers and agree that it</p>	<p>Convergent validity: Children's scores on COR were found to correlate highly with the relevant subscales of</p>	<p>Multidimensional Rasch models to evaluate the structural and substantive aspect</p>	<p>Developmental validity: Barghaus & Fantuzzo (2014) using IRT revealed some items</p>	<p>NA</p> <p>Widely used measure in the US. Included in a review by Halle et al., (2011) as one of 8 formative assessment</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>day writing brief notes or “anecdotes” that objectively describe significant episodes of children’s activities. The anecdotes are classified and scored according to various COR categories, items, and levels, providing a comprehensive portrait of each child’s developmental gains and the progress of the group as a whole.</p> <p>Latest version: 2019</p> <p>Source: https://highscope.org/cor-advantage/</p> <p>Age range: 0:0–6:0 years</p> <p>Cost: Request quote</p> <p>Number of items: 34 items across eight areas or categories comparable with the content of national and state standards, plus two items for English Language Learning (ELL).</p> <p>Time: Unspecified</p>	<p>and usually lasts 2–3 months, up to 3/year.</p> <p>Online or IT-based mode: Yes. COR is administered mainly as a web-based online assessment (including a tablet version).</p> <p>Standardised scoring: NA</p> <p>Training: COR training includes website navigation and interface functions, interpretation of items/domains, how to write and score objective anecdotes, as well as tips for collecting anecdotes throughout the day.</p> <p>Who administers the tool and receives feedback: Educators. Parents can also contribute anecdotes and view their child’s uploaded photos and work.</p> <p>Details included in feedback: The website generates a variety of reports to help teachers plan instruction, including growth profiles for each child and reports on group progress. Group reports can be aggregated at various levels (e.g., state, region, grantee, program, site, and classroom) and fully</p>	<p>adults, building relationships with other children, community.</p> <p>2. Connected with and contribute to the world <i>Science and Technology:</i> observing and classifying, experimenting, predicting and drawing conclusions, natural and physical world, tools and technology. <i>Social studies:</i> knowledge of self and others, geography, history.</p> <p>3. Sense of well-being <i>Social and emotional development:</i> emotions, building relationships with adults, building relationships with other children, community, conflict resolution. <i>Physical Development and Health:</i> fine and gross motor skill, personal care and healthy behaviour.</p> <p>4. Confident and involved learners <i>Approaches to learning:</i> initiative,</p>	<p>accurately assesses key development domains (Wakabayashi et al., 2019; Waterman et al., 2012).</p> <p>Inter-rater reliability: After appropriate training, teachers attained high levels of agreement with experts across all children and items (Wakabayashi et al., 2019; Waterman et al., 2012).</p>	<p>Woodcock-Johnson III Tests of Achievement and Social Skills Improvement System (Wakabayashi et al., 2019).</p> <p>Divergent validity: There are disagreements regarding the factor structure due to high correlation between domains. Barghaus & Fantuzzo (2014) present a 4-factor model of social engagement, cognitive skills, coordinated movement and scientific process skills. Possibility that a second order factor common to all items.</p>	<p>of construct validity support the theorized eight dimensions underlying the COR Advantage and suggests that scoring rubrics function as the instrument developers intended (i.e., as an eight-level rating scale) (Wakabayashi et al., 2019).</p>	<p>have skill points that do not indicate a developmental progression. For Head Start teachers applying the COR average assessor variance is 27.6%, indicating that only 70–80% of score variation is child centred (Waterman et al., 2012). No investigation of differential item functioning across various contexts.</p> <p>Preschool scores from the cognitive dimension (which includes literacy and numeracy) predict DIBELS letter naming fluency in kindergarten and DIBELS oral reading fluency in grade 1 (~11-14% of variance accounted for), (Singer, 2009).</p>	<p>tools meeting criteria that included a) covered three or more of the domains of the Head Start Child Outcomes Framework, (b) had some evidence base, and (c) were accessible for general use.</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
	<p>integrate with the reporting requirement of the Office of Special Education Programs and Head Start. There is a library of developmentally appropriate activities aligned with each COR level. These resources are available for users as ideas to support children's development.</p> <p>Suitable for diverse populations: Yes. COR is designed to capture developmental trajectories of all children.</p>	<p>problem solving, reflection. <i>Mathematics:</i> number and counting, shapes and spatial awareness, measurement, patterns, data analysis. <i>Creative arts:</i> art, music, movement, pretend play.</p> <p>5. Effective communication <i>Language, literacy, and communication:</i> speaking, listening and communication, phonological awareness, alphabet knowledge, reading, book enjoyment and knowledge, writing.</p>					
<p>Desired Results Developmental Profile (DRDP)</p> <p>Formative assessment instrument developed by the California Department of Education for young children and their families used to inform instruction and program development.</p> <p>All assessments can be in the child's first language. There are 4 additional assessments of English Language and Literacy for children who do not have</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Yes</p> <p>Standardised scoring: Each item is rated on a continuum of skill development (not yet exploring, exploring, developing, building, and integrating. After rating the initial level teachers then rate the child as emerging if the child is beginning to show some skills from the next level (considered a half point on the measure). Skill</p>	<p>1. Sense of identity <i>Social and Emotional domains:</i> identity of self in relation to others, relationships and social interactions with familiar adults, relationships and interactions with peers.</p> <p>2. Connected with a contribute to the world <i>History and social science domain:</i> sense of time and place, ecology,</p>	<p>Internal reliability: Reliability indices ranged from 0.73 to 0.99, indicating that DRDP (2015) domains and sub-domains all had adequate score reliability. The separation reliability indices of 0.99 indicated that the developmental levels within each DRDP (2015) grouping of domains were highly distinct.</p> <p>Inter-rater reliability: 31 unique assessor</p>	<p>Convergent validity: Alignment with the child completed adaptive assessment (language and math concepts) correlations were low to moderate (Moiduddin et al., 2014). Assessments were collected for 126 preschool-aged children. DRDP results were correlated with Expressive One Word Picture Vocabulary Test, Receptive One Word Picture Vocabulary Test, Woodcock-Johnson III Achievement tests, and Preschool and</p>	<p>Factor analysis supported a 5-factor structure of the tool at all three time points (self-awareness and identity, mathematics, social skills, language and literacy, general cognitive; Nguyen et al., 2019).</p>	<p>Developmental and subgroup validity: The peak of the intraclass correlations (ICC) for each level was expected to be above a probability rating of 0.5. The research team determined that all of the DRDP measures – and rating categories – demonstrated adequate functioning. The ICCs were ordered, and the category peaks showed distinct</p>	<p>Krause, 2016 (T) – Interviews with 20 educators from Head Start and state preschools. 60% of teachers noted that reflecting on DRDP results is challenging due to time constraints and requires allocated time off the floor with children. 20% wished there were fewer items (suggesting that some items were asking about the same thing). Teachers find the WestEd website helpful in implementing DRDP and useful program-specific DRDP resources are provided by the district or school, e.g., observation recording templates, i-pads.</p> <p>Moiduddin et al., 2014 (S/T) – Interviews with 7 teachers and 2 supervisors. See 'Developmental and subgroup validity' for reported psychometrics. Most</p>

Tool			Psychometrics			Evidence	
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>English as their first language.</p> <p>Latest version: 2015</p> <p>Source: www.desiredresults.us/drpd-forms</p> <p>Age range: 3:0–5:0 years (kindergarten assessment available)</p> <p>Cost: Scoring rubric free, unclear if online tools are commercialised.</p> <p>Number of items: Varies depending on the comprehensiveness of the required assessment (comprehensive, fundamental, essential and snapshot). Max = 52 (comprehensive preschool version). Min = 17 (snapshot kindergarten version).</p> <p>Time: Variable</p>	<p>points are then averaged across all items within each category to create domain scores and across domains to create a total score.</p> <p>Training: At least some school districts in the US require educators to undertake training for DRDP (Krause, 2016).</p> <p>Who administers the tool and receives feedback: Educators, can include family observations.</p> <p>Details included in feedback: Unspecified</p> <p>Suitable for diverse populations: Specific consideration for children who are dual language learners. Developed with the goal of ensuring that all children have the opportunity to demonstrate their knowledge and skills. To enable access to the assessment for diverse populations, the principles of Universal Design were followed. Categories of adaptations to support specific groups are also provided.</p>	<p>conflict negotiation, responsible conduct.</p> <p><i>Visual and Performance Arts:</i> visual art, music, drama, dance.</p> <p>3. Sense of well-being <i>Social and Emotional domains:</i> identity of self in relation to others, social and emotional understanding, relationships and social interactions with familiar adults, relationships and interactions with peers, symbolic and sociodramatic play.</p> <p><i>Physical Development – Health:</i> perceptual-motor skills and movement concepts, gross locomotor movement skills, gross motor manipulative skills, fine motor manipulative skills, active physical play, nutrition, safety, and personal care routines (hygiene, feeding, dressing).</p> <p>4. Confident and involved learners <i>Approaches to Learning – Self-regulation:</i> attention</p>	<p>pairs completed the DRDP assessment for 1 to 4 children per pair. Interrater agreement percentages were calculated for both exact agreement (results ranged from 48% to 81%) and agreement within one rating level ranged from 83 to 98%. For preschool-aged children interrater agreement ranged from 50 to 75%, and agreement within one rating level from 84 to 97%. Exact agreement for domain- scaled ratings ranged from 92 to 97% for preschool-aged children (DRDP Technical Report, 2015).</p>	<p>Kindergarten Behaviour Scale (DRDP Technical Report, 2015).</p>		<p>separation across the distribution of ability. More than 92% of children progressed on each of the child outcomes, demonstrating sufficient sensitivity.</p> <p>39-60% of children advanced by at least one developmental level between two measurement points, and no measure was overly easy (>90% children advancing one level between two time points) or overly hard (<10% advancing). This was also true for children with mild and severe limitations.</p> <p>Some support that the DRDP is sensitive to children's progress over the year. Most children moved forward in DRDP rating, but 5 to 9% children received a lower rating in the spring relative to the fall (Moiduddin et al., 2014).</p> <p>Items assessing concepts about print, emergent writing, and phonological awareness did not</p>	<p>teachers combined direct assessment with observations and ongoing documentation to compile a portfolio, which was then used to complete the DRDP (either all at once, or focusing on a few measures across multiple children). A few teachers explicitly assessed children's skills by posing specific questions in one-on-one situations or setting up tasks to assess a specific set of skills. All teachers said they complete the DRDP after school hours and/or at home, taking 20-40 minutes per child. Results are manually summarised in the classroom-level summary sheet. DRDP results were used to select objectives, plan whole-class activities, and sometimes to form small groups; only a few teachers used DRDP results for individualised instruction. Most teachers received no formal training in DRDP, despite average use time of 8.4 years. Supervisors provided support to teachers on DRDP interpretation, goal setting and lesson planning at monthly or one-on-one meetings. Teachers report not having time to collaborate with colleagues outside their own classroom or to have longer discussions with parents; they also have limited communication with K/Yr1 teachers. DRDP is perceived as useful for lesson planning and organising levelled groups according to children's strengths and needs, but some teachers said that the results lack nuance and therefore are not useful, and that the teacher-generated assessment is highly subjective.</p> <p>Piper et al., 2013 (T) – interviews with 9 teachers at a university ECE centre. Photos and photo sequences are the</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
		<p>maintenance, engagement and persistence, curiosity and initiative, self-comforting, self-control of feelings and behaviour, imitation, shared use of space and materials.</p> <p><i>Cognition and number:</i> spatial relationships, classification, number sense of quantity, measurement, patterning, shapes.</p> <p>5. Effective communication <i>Language and literacy development:</i> understanding of receptive language, responsiveness to language, communication and use of expressive language, reciprocal communication and conversation, interest in literacy, comprehension of age appropriate text, concepts about print, phonological awareness, letter and word knowledge, emergent writing</p>				function equivalently for dual language and non-dual language children (Nguyen et al., 2019), however this study only included low income children who were also majority Hispanic and DLL.	most common form of evidence, followed by samples of child's work. Teachers focus on a smaller (3-8) group of children to observe better and focus on critical episodes of activity and try to seek authentic observation in line with Reggio Emilia approach. Observations are typically written down soon after they take place, however many teachers rely on note paper/post-its before filling out an observation sheet (which is often done at home/after hours). More observations than necessary are collected, so ECEs selected which ones are best to include. Teachers may share observations and work together to complete the DRDP, which creates inconsistencies in the documentation format/style. Teachers and school administrators calculate developmental progress for individuals, classes, age groups, and the school as a whole. They use these calculations to plan and tailor curriculum. Aggregating ratings from these paper forms and deriving new goals based on this data is a challenge for teachers and school administrators.
Early Learning Scale	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Yes</p>	2. Connected with a contribute to the world Math/science domain.	Inter-rater reliability ranged from .71 to .77 (good) for teachers, and from .91 to .98 (excellent) for trainers	Concurrent validity: ELS had moderate correlations with literacy and mathematics scales, ranging from .39 to .46	NA	NA	NA
A systematic assessment designed to assess progress							

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>towards learning standards such as Head Start Child Outcomes Framework and New Jersey Learning Expectations. Includes Math/Science, Social-Emotional/Social Studies, and Language and Literacy domains.</p> <p>Latest version: 2010</p> <p>Source: www.myelsonline.com</p> <p>Age range: "Pre-K"</p> <p>Cost: Annual subscription \$9.95–15.95 per child</p> <p>Number of items: 10</p> <p>Time: <5 min</p>	<p>Standardised scoring: Yes</p> <p>Training: Extensive training available through the provider (\$3500 for 2 days + reliability assessment after first tool use period)</p> <p>Who administers the tool and receives feedback: Educators</p> <p>Details included in feedback: Ideas for teaching and documenting and a list of resources for further reading for each item.</p> <p>Suitable for diverse populations: Unspecified</p>	<p>3. Sense of wellbeing Social-Emotional/Social Studies domain.</p> <p>5. Effective communication Language and Literacy domain.</p>	<p>(Riley-Ayers et al., 2016).</p> <p>Internal reliability: Cronbach's alpha of .91, indicating high internal reliability (Riley-Ayers et al., 2016).</p>	<p>on whole instrument comparisons (Riley-Ayers et al., 2016).</p>			
<p>Early Literacy Individual Growth and Development Indicators (EL-IGDIs)</p> <p>Constructed with Item Response Theory, offering a precise method of locating an individual on an ability-based scale. EL-IDGs were developed to inform decisions about whether children are demonstrating adequate levels of performance given the general level of instruction (Tier 1), or if their performance indicates a need for more intense levels of instruction (Tier 2 or Tier</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Administered on paper, online system to store scores is available.</p> <p>Standardised scoring: Roseth et al., (2012) aimed to produce age-based norms that practitioners can use to evaluate growth and status at any given time.</p> <p>Training: Tool administrator should be trained in the use of test materials, but the tool and the website are designed with ease of</p>	<p>5. Effective communication Picture Naming (Oral Language) Rhyming (Phonological Awareness) Sound Identification (Alphabet Knowledge) 'Which One Doesn't Belong?' (Comprehension) Alliteration (Phonological Awareness)</p>	<p>Test-retest reliability: McConnell et al., (2012) report test-retest reliability as .93–.97, which is considered excellent.</p>	<p>Concurrent validity: Roseth et al., (2012) found medium to large correlations between EL-IDGI domains and standardised measures (PPVT-4, TOPEL, TOPEL-PA, TOPEL-PK, CELF-Pre-2), ranging from .52 to .71.</p>	NA	<p>Developmental and subgroup validity: Roseth et al., (2012) report sensitivity of .71–.77 and specificity of .57–.69, implying that we cannot assume linear growth in EL-IDGI data across the entire developmental period of the scale (30–66m), especially for rhyming and alliteration. Rhyming and alliteration measures may have a floor effect, or the skills themselves require greater maturation than the</p>	<p>Gettinger & Stoiber, 2012 (S) – teachers of 62 students received PD for 4 months and 62 were a control group. PD included EMERGE literacy coaching, using data from FA and using it to inform instruction in small groups. Students in the intervention group showed a significant improvement on measures of vocabulary ($\eta^2 = .18$), rhyme awareness and alphabet knowledge ($\eta^2 = .05$), Head Start outcomes (Story and Print Concepts Task; $\eta^2 = .10$), and early literacy skills ($\eta^2 = .10$). When fall assessment scores were controlled for, there was no main effect for level of individualised intervention (high, medium, low). None of the treatment x level interactions were significant. Therefore, treatment did better than control, but there was no difference for students at different performance levels.</p>

Tool						Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
Description		EYLF Outcomes	Reliability				
<p>3). Includes comprehension and alphabet knowledge, picture naming, rhyming, and alliteration measures.</p> <p>Latest version: ongoing updates</p> <p>Source: https://www.myigdis.com/preschool-assessments/early-literacy-assessments/#1460350206401-de3ecdce-928f</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Request quote</p> <p>Number of items: 5 literacy domains each with 15 test items. Assessment is time limited so not all items will be completed.</p> <p>Time: 10 min</p>	<p>administration and fidelity in mind. The coordinator who oversees EL-IGDIs use and interpretation should be a professional in special education, early childhood education, psychology, speech and language, school nursing, or another closely related area. A range of PD opportunities are provided focusing on tool administration and more broadly on assessment and using MTSS/RtI as a decision-making framework.</p> <p>Who administers the tool and receives feedback: Educators</p> <p>Details included in feedback: Overall score</p> <p>Suitable for diverse populations: Most of the studies are focused on typical children. One earlier study suggesting significantly different intercepts and slopes for children with speech or language difficulties, children living in poverty, and dual-language learners (Missall et al., 2006). Items were not designed to avoid cultural bias, but the diversity of items sampled within the timed</p>					<p>picture-naming measure.</p> <p>Evidence from Roseth et al., (2012) suggests the tool might not be sensitive or that a one-time assessment might not be appropriate given that there are age-related differences in growth rate.</p> <p>Note that there is ongoing funded work (from the Institute of Educational Sciences, US) to continue to refine these tools).</p>	<p>Authors note that future research should add a group that receives Tier 1 intervention (PD) but no progress-monitoring to differentiate their effects. Also note that the progress-monitoring probes were tightly aligned with curriculum and may not capture broader literacy growth.</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
	administration period may statistically control for bias experienced by any one child/group.						
<p>Early Numeracy - IDGIs (also called Preschool Numeracy Indicators)</p> <p>IDGIs Numeracy is a collection of progress-monitoring measures, covering oral counting, quantity comparison, number naming, and 1-to-1 correspondence counting. The tool is intended to measure developmental gains and individualise instruction.</p> <p>Latest version: 2013</p> <p>Sources: www.myigdis.com/preschool-assessments/</p> <p>Age range: 3:0–6:0 years</p> <p>Cost: Request quote</p> <p>Number of items: Exact number not specified. Discontinue rule applies.</p> <p>Time: 10 min</p>	<p>Intended cycle frequency: Up to 1/month</p> <p>Online or IT-based mode: Administered on paper, online system to store scores is available.</p> <p>Standardised scoring: Norm referenced growth standards available.</p> <p>Training: Same as EL-IDGIs.</p> <p>Who administers the tool and receives feedback: Educators, reports can be shared with parents.</p> <p>Details included in feedback: Overall score</p> <p>Suitable for diverse populations: Standardised for English speakers only.</p>	<p>4. Confident and involved learners</p> <p>Oral Counting, Quantity Comparison, Number Naming, 1-to-1 Correspondence Counting.</p>	<p>Test-retest reliability: According to the publisher technical report (Hojnoski & Floyd, 2013), test-retest reliability is .71–.88, which is considered good.</p>	<p>Concurrent validity: Concurrent validity with Woodcock-Johnson III Applied Problems, Bracken Basic Concept Scale – Revised, and Test of Early Mathematics Ability – Third Edition ranges between .60 and .75 (Hojnoski & Floyd, 2013).</p>	NA	<p>Developmental and subgroup validity: IDGIs-Numeracy domains were found to be sensitive enough for predicting achievement below 40th percentile, but specificity was poor – 33–54% of typically achieving children were predicted to be in the lowest 40th percentile. For prediction of bottom 25th percentile, sensitivity is lower, but specificity is higher (Laracy et al., 2016). This suggests the tool may not be sufficient for differentiating between younger or developmentally delayed children, at least as a one-time assessment.</p>	NA
<p>Early Numeracy Scales (ENS)</p> <p>Brief measures to identify individual needs and measure progress, in the</p>	<p>Intended cycle frequency: Unspecified</p> <p>Online or IT-based mode: No</p>	<p>4. Confident and involved learners</p> <p>A variety of mathematical sub-domains. 1-to-1 counting, cardinality,</p>	<p>Test-retest reliability: All tasks had significant small to large correlations ($r = .22-.72$) with the same tasks given a year</p>	<p>Concurrent validity: ENS had significant small to medium correlations with all Woodcock-Johnson III Applied Problems (.35–</p>	NA	<p>Developmental and subgroup validity: No differential item functioning was detected for any items</p>	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>domains of 1-to-1 counting, cardinality, counting subsets, subitizing, number comparison, set comparison, number order, numeral identification, set-to-numerals, story problems, number combinations, and verbal counting.</p> <p>Latest version: 2015</p> <p>Source: Purpura & Lonigan (2015)</p> <p>Age range: 3:0–6:0 years</p> <p>Cost: Unspecified</p> <p>Number of items: 68</p> <p>Time: Each measure takes <5 min to administer (2-3min most common).</p>	<p>Standardised scoring: Yes</p> <p>Training: Unspecified</p> <p>Who administers the tool and receives feedback: Research assistants, educators.</p> <p>Details included in feedback: A score for each domain.</p> <p>Suitable for diverse populations: The 393 children who completed the assessment were about evenly split by sex (51.7% female) and were approximately representative of the demographics of the local area (55.7% White, 33.8% African American, and 10.5% other race/ethnicity). Children ranged in age from 3.13 to 5.98 years (M = 4.75 years, SD = 0.75 years), were primarily English speaking, and had no known developmental disorders.</p>	counting subsets, subitizing, number comparison, set comparison, number order, numeral identification, set-to-numerals, story problems, number combinations, and verbal counting	later, except cardinality (in all children), and number combinations (in younger children). This was likely caused by restricted range of these tasks (Purpura & Lonigan, 2015).	.66) and Calculations tasks (.22–.55) for both older and younger children. (Purpura & Lonigan, 2015). Predictive validity: All ENS tasks had significant small to large correlations with Woodcock-Johnson III Applied Problems (.20–.70) and Calculation (.20–.54). The only exception was cardinality in younger children, which did not have a significant correlation with Calculation a year later. (Purpura & Lonigan, 2015).		<p>based on sex or ethnicity.</p> <p>It is not intended that a teacher would utilise all 12 assessment tasks at the same time for a child. The tasks can be used to measure the progress when a child is learning concepts from the relevant domain.</p> <p>Note: "Older children" are the ones who started kindergarten in Year 2 of the study.</p>	
<p>Google Forms based on Behaviour Incident Reporting System (BIRS)</p> <p>The 3 tools are combined to make a cloud-based knowledge management to facilitate decision making.</p>	<p>Intended cycle frequency: NA</p> <p>Online or IT-based mode: Yes</p> <p>Standardised scoring: No</p> <p>Training: No, but authors report the pilot team planning a PD day</p>	NA	NA	NA	NA	NA	Johnson (2017) reports some early results from the pilot. System was adopted in a Midwestern USA district, with 36 educators. Teachers were very satisfied with the system bypassing paper forms entirely. Links were made available either from home screen of classroom iPads, or from QR codes within the classroom that linked to that classroom's Google form. However, there were difficulties obtaining the initial

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>BIRS was made into a google form, the data is automatically put into google sheets for easy summarising, and can also be copy pasted into Excel spreadsheets that's set up to create useful graphs and summaries to facilitate decision-making. The district where this is being piloted uses the Pyramid Model, so the tools are aligned with it.</p> <p>Latest version: 2017</p> <p>Source: Johnson, 2017</p> <p>Excel workbook: https://1drv.ms/x/s!Av4jJ8fn9JqdhI-8E9mEvDi-wHI3Q</p> <p>Template for google sheets/forms: https://docs.google.com/spreadsheets/d/1GHwJIEOc3ECO1qNbsIJSMiFotyU3YiW_4vo6Ha8jbZ8/edit?usp=sharing</p> <p>Age range: NA</p> <p>Cost: Not a commercial tool</p> <p>Number of items: NA</p> <p>Time: NA</p>	<p>to let teachers practice interpreting the summary sheet and exploring action items based on each interpretation.</p> <p>Who administers the tool and receives feedback: Educators; parents can receive feedback if necessary.</p> <p>Details included in feedback: The Excel Workbook provides various summaries, such as the number of incidents over time broken down by activity type, as well as calculated risk/probability of future incidents broken down by demographics. The summaries can help identify 'difficult' activities/groups of children and facilitate decision-making. Can also see individual summaries by child and use that for progress-monitoring.</p> <p>Suitable for diverse populations: NA</p>						<p>enrolment information as it's stored in multiple places.</p> <p>An item was added to the Google Form (whether assistance was needed during the incident) as a result of using the system and identifying a need for additional source of information.</p>
Individualized Classroom Assessment Scoring System (inCLASS)	<p>Intended cycle frequency: Unspecified</p> <p>Online or IT-based mode: No</p>	3. Sense of wellbeing Positive Engagement with the Teacher; Teacher Communication;	Inter-rater reliability: Coders were within one point of each other's scores 87% of the time (with a range of 71% to 99% across	Concurrent validity: InCLASS scores had significant small to medium correlations with teacher ratings on other established	EFA with oblique rotation strongly supported a 3-factor model (Teacher Interactions, Peer Interactions, and	Developmental and subgroup validity: InCLASS scores did not significantly differ by sex in an earlier study (Downer et al.,	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>inCLASS is an observation tool that targets children's interactions in preschool classrooms with teachers, peers, and tasks. The nine dimensions include: positive engagement with the teacher, teacher communication, teacher conflict, peer sociability, peer assertiveness, peer communication, peer conflict, engagement within tasks, and self-reliance.</p> <p>Latest version: 2018</p> <p>Source: www.inclassobservation.com</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Commercial product but cost unclear</p> <p>Number of items: There are 9 dimensions, and the child receives a global score on a 7-point scale for each dimension based on the observation of specific behavioural markers, which are developmentally graded.</p> <p>Time: 15min/observation cycle</p>	<p>Standardised scoring: Yes</p> <p>Training: 2-day training program available from the provider. Observers in Downer et al., (2010) received two days of training and had to score within 1 point of the master-code on 80% of their scores to be deemed reliable and ready for live data collection.</p> <p>Who administers the tool and receives feedback: Research assistants, educators</p> <p>Details included in feedback: Sub-domain scores</p> <p>Suitable for diverse populations: The tool has been validated on a gender-balanced, ethnically and socioeconomically diverse sample (Downer et al., 2010). Bohlmann et al., (2018) found that inCLASS works adequately for children of different ethnicities and socioeconomic status, but the results for gender were mixed.</p>	<p>Teacher Conflict; Peer Sociability; Peer Assertiveness; Peer Communication; Peer Conflict; Engagement within Tasks; Self-Reliance;</p> <p>4. Confident and involved learners Engagement within tasks; Self-reliance;</p> <p>5. Effective communication Teacher and peer communication domains</p>	<p>the nine inCLASS dimensions). An intraclass correlation was also calculated across all dimensions and reached 0.84. Intraclass correlations at the domain and dimension levels ranged from moderate to excellent (0.42 – 0.83) (Downer et al., 2010).</p>	<p>measures (Academic Rating Scale – Literacy and Language, Student Teacher Relationship Scale, Teacher-Child Rating Scale, and California Preschool Social Competency Scales) (Downer et al., 2010).</p> <p>Discriminant validity: The pattern of significant results and the direction of the correlation generally lined up with what was expected for all four domains of inCLASS. Unexpectedly, Task Orientation was significantly correlated with social communication, assertiveness, and social skills. The authors suggest that children may use these traits when approaching learning tasks too (Downer et al., 2010).</p>	<p>Conflict Interactions), with some evidence for a fourth factor (Task Orientation). Item loadings ranged from .68 to .94, and all cross-loading were below .53. Task Orientation was kept as a fourth factor, and this 4-factor model accounted for 85.71% of the variance in inCLASS observations (Downer et al., 2010).</p> <p>Bohlmann et al., (2018) conducted a CFA using robust maximum likelihood, which supported the 4-factor model with a more diverse sample.</p>	<p>2010), but more complex analyses yielded mixed results (Bohlmann et al., 2019). Bohlmann et al., (2019) found the data to fit the same 4-factor model for poor children as compared to non-poor, as well as for African American, Hispanic, and White children. Age was positively correlated with Peer Interaction scores ($r = .48, p < .001$) and Task Orientation scores ($r = .22, p < .01$), but not Teacher Interaction or Conflict Interaction scores. This suggests that the scoring may be sufficiently sensitive in peer interaction and task orientation domains.</p>	
<p>Learning Stories</p> <p>A narrative account of an incident that has taken</p>	<p>Intended cycle frequency: NA</p> <p>Online or IT-based mode: Yes</p>	<p>Learning Stories can be linked with any of the five EYLF learning outcome areas by the</p>	<p>All observations are open to interpretation in different ways and are meant to be subjective. Hence,</p>	<p>Lack of evidence on validity or credibility; problems in trying to use LS to show changes in children's learning over</p>	<p>NA</p>	<p>Unclear how the data resulting from the use of Learning Stories would be sensitive to small developmental</p>	<p>Goodine, 2013 (T) – qualitative study with 36 children and their teachers in Canada. The centre used developmental checklists to assess kindergarten readiness and portfolios to collect all</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>place. The story can involve individuals or groups. The teacher reflects on the story using learning dispositions to analyse the learning that has taken place. The learning dispositions are: taking an interest, being involved, persisting with difficulty, expressing an idea or a feeling and taking responsibility. Often photographs are used for illustration and to make the story accessible to the child/children, which enhances its potential for revisiting.</p> <p>Latest version: NA</p> <p>Source: Originally developed by Carr (1998, 2001). Updated publication Carr & Lee (2012).</p> <p>Age range: NA</p> <p>Cost: Varies depending on source. Possibility that providers could develop their own tool. For an online template that links to EYLF, costs vary. StoryPark = 99c per month per child; EarlyWorks = \$1.45 per month per child.</p> <p>Number of items: NA</p> <p>Time: NA</p>	<p>Standardised scoring: No</p> <p>Training: NA</p> <p>Who administers the tool and receives feedback: Educators, but children are also involved in the documentation of the story (intended to promote capacity for self-reflection). Feedback can be received by educators, the child, family, and other members of the teaching team.</p> <p>Carr describes the writing of a learning story as a dialogic process in which the documenting educator discusses observations and their interpretation of it with the child and other colleagues.</p> <p>Details included in feedback: NA</p> <p>Suitable for diverse populations: Theoretically should be suitable for any population, although educator's biases could lead to unfavourable interpretations of events and learning for some children. The lack of systematicity in the approach may lead to overlooked development</p>	educators writing the stories.	evaluating reliability is not possible. Kalliala & Pramling Samuelsson (2014) comment that pedagogical documentation 'is prioritized and censored before it is written'.	time (Blaiklock, 2013). May not be possible to provide valid evidence about how children are progressing in key learning areas.		changes seen during the course of one year (Blaiklock, 2013).	<p>child-related information. Portfolios were not shared with parents or children, and in some cases did not accurately reflect the child's strengths and areas of concern. During the study, portfolios were moved to the playroom and made accessible to both ECEs and children. Individual cameras and journals provided by the researchers made the data collection more streamlined. All ECEs commented on information-overload, and lack of support and time for planning, reflection, and documentation; ECEs relied on the children's rest and outdoor play periods to collaborate, reflect and document data.</p> <p>During the course of the study, the ECEs became aware of their undocumented assessment practices. 2/3 ECEs recalled receiving Learning Stories-related training in their classes. For all 3 ECEs, there was a focus on identifying developmental progression, but not using assessment to inform planning or documenting child's learning/capabilities. Different ECEs subjectively focused on different domains of learning, both in the recorded observations and the activities they planned, e.g. creative and emotion domains were assessed less often than gross and fine motor, or social skills.</p> <p>Lack of resources was identified by Program Administrators as the main barrier to use of rubric assessments. Resources were not well developed and not much used for developmental checklist and portfolio. Narrative progress reports are not used due to lack of training. Observations and parent feedback were reportedly used when they were concerned about a child's development. Most common supports for</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
	in some areas for any child (Blaiklock, 2013).						<p>assessment practices were team teaching and providing access to information technology (every other week most of the time). Paid time away from children to interpret and document learning was provided on average once a month.</p> <p>Hooker, 2017 (T) – collective case study with 6 families and their teachers in NZ, comparing paper-based and e-portfolios. E-portfolios resulted in more consistent and frequent entries, with an average of 1/week, used more multimedia and involved children in writing and selection of media. Teachers felt more motivated to make new entries because they knew the entry would be seen by parents in a shorter timeframe. The electronic format was reported to save time as compared to paper-based portfolios, allowed to make connections between learning and experiences more often, and allowed for more collaboration between teachers.</p> <p>Hooker, 2019 (T) – nested case studies with 6 teachers, 6 parents and 6 children in NZ. This findings are generally in line with Hooker (2017). After introduction of e-portfolios, children became more involved in conversations about learning and were more supported to continue their learning at home. E-portfolios allowed to keep videos of children, which were informative for the adults and enjoyable for children.</p> <p>Loggenberg, 2011 (T) – 25 completed questionnaires from ECE centres and kindergartens, as well as 3 from ECE teacher programs in NZ. Educators most commonly used (and providers encouraged using) learning stories</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
							<p>(80%) and checklists were least common (8.8%). Teachers did not use checklists primarily because they were considered not age appropriate (24%) and not for ECE teachers to use (24%). Some commented that LS are time-consuming and do not allow them to effectively identify early obstacles experienced by learners. 2/3 education providers commented that LS often happen without actual assessment/not used for ongoing assessment, and that the lack of developmental observation skills detracts from LS value. Unlike educators, all 3 providers said that checklists are a useful tool.</p> <p>Teachers reported using information from assessment to improve the achievement of individual learners and groups (92%), to inform strategic planning and service self-review (88%) and evaluate the success of the service's curriculum and teaching programmes (84%). All 3 ECE providers responded similarly. Not all areas were assessed equally commonly, because teachers needed training in the assessment area, do not consider it suitable for EC educators, or have not heard of the area before. Literacy and numeracy were rarely assessed (35-36%) despite high endorsement from ECE providers (80-86%).</p> <p>Niles, 2016 (T) – interviews with 12 teachers in NZ. Some teachers struggled with documenting children's perceived area of need and children setting their own goals / finding solutions (key feature of Kei Tua o te Pae), as well as deciding what to record and whether a child or a group should be described. Some evidence was collected in response to a</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
							<p>parent's concerns, as means of accountability. Children's weaker areas were discussed in meetings and their development supported, but teachers were hesitant to record these in LS because they are hard to write in a positive light. Meetings were regular but mostly taken up by housekeeping issues; this was mitigated by teachers reading children's portfolios in pairs as means to collaborate. Teachers were concerned about time pressure (~2h/wk non-contact time is needed), the 'right' way to write LS, and how precise the set goals should be. Overall teachers were supportive of LS and sought to improve their practice.</p> <p>Nyland & Alfayez, 2012 (T) – interviews with 3 ECE lecturers and 3 students in AU. All expressed enthusiasm about LS and its benefits in providing a format for recording and sharing observations, and a way to emphasise participation of children, educators, families, and the community. Two students said LS are best for small groups and take a long time. Two lecturers believed that the to use Learning Stories in an effective and efficient way the practitioner needs experience, practice, and time to reflect on available information.</p> <p>Zhang, 2017 (T) – interviews with managers, educators, and parents (35 total) in NZ. Zhang highlights differences between the 3 groups and suggests there are tensions between everyday practices and quality assurance authorities. For example, practitioners and parents highlighted that milestones of learning and development should be captured by a range of assessment tools. The New Zealand Education</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
							Review Office's focus on predominant use of LS make this difficult because comparison across LS is not possible given the anecdotal nature and limited frequency. Zhang highlights a conflict between the Education Review Office's stance on LS as the preferred assessment for learning versus the views of practitioners and parents that there should be multiple forms of assessment.
<p>Letter-Sound Short Forms</p> <p>A set of formative assessments of letter-sound knowledge developed specifically for the preschool period.</p> <p>Latest version: 2016</p> <p>Source: Piasta et al., (2016)</p> <p>https://earlychildhood.ehe.osu.edu/files/2016/04/Letter-Sound-Short-Form-Assessment1.pdf</p> <p>Age range: 3:0–6:0 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: 8</p> <p>Time: 3–5min</p>	<p>Intended cycle frequency: 3–4/year</p> <p>Online or IT-based mode: No</p> <p>Standardised scoring: Yes</p> <p>Training: No</p> <p>Who administers the tool and receives feedback: Research assistants, intended for educators.</p> <p>Details included in feedback: A scaled or a sum score.</p> <p>Suitable for diverse populations: NA</p>	<p>5. Effective communication</p> <p>Alphabet knowledge letter-sound knowledge</p>	<p>Parallel forms reliability:</p> <p>Reliability for all versions of the short forms was high: $\rho = .92-.93$ for three-form version, and $\rho = .89-.91$ for the four-form version. There were no differences between development and validation datasets ($\rho = .849$) and the full range of letter-sound knowledge (0-26) was demonstrated (Piasta et al., 2016).</p>	NA	<p>Exploratory factor analysis strongly supported a unidimensional model (first eigenvalue = 21.016, other eigenvalues ranging from .693 to .012). Confirmatory factor analysis on the validation dataset also supported a one-factor model (CFI = .993, TLI = .999, RMSEA = .040), (Piasta et al., 2016).</p>	NA	NA
<p>Preschool Early Literacy Indicators (PELI)</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Pen-and-paper form and storybooks, has</p>	<p>5. Effective communication</p> <p>Literacy alphabet knowledge, phonological awareness,</p>	<p>Inter-rater reliability</p> <p>ranged between .91 and .99 on five alternate forms for all subtests (Aguayo & Kaminski, 2012),</p>	<p>Concurrent validity:</p> <p>PELI composite score had significant medium correlations with all criterion measures (.42–.62). The alphabet</p>	NA	<p>Developmental and subgroup validity:</p> <p>Kaminski et al., (2018) examined the differences in growth in PELI scores</p>	<p>Greenwood et al., 2017 – see CIRCLE (former CPALLS+STEM) Evidence section.</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>The PELI assessment was designed for screening and progress monitoring of preschool children's acquisition of early literacy and language skills. The PELI is comprised of four subtests: alphabet knowledge, phonological awareness, vocabulary (oral language), and listening comprehension.</p> <p>Latest version: 2012</p> <p>Source: https://acadiencelearning.org/peli.html</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Unspecified</p> <p>Number of items: Unspecified</p> <p>Time: 15 minutes, can be over multiple occasions</p>	<p>online database entry system.</p> <p>Standardised scoring: Yes</p> <p>Training: Basic administration training is needed.</p> <p>Who administers the tool and receives feedback: Educators</p> <p>Details included in feedback: Subscale and composite scores.</p> <p>Suitable for diverse populations: NA</p>	<p>vocabulary (oral language), and listening comprehension</p>	<p>which is considered excellent.</p> <p>Parallel form reliability: Parallel form reliability for three PELI forms ranged from .87 to .96 in one study (Kaminski et al., 2014) and ranged between .89 and .94 for five alternate PELI forms in another (Aguayo & Kaminski, 2012). Individual subtests for the three forms had a greater reliability range of .58-.98. (Kaminski et al., 2014).</p>	<p>knowledge subtest significantly correlated with criterion measures in some studies (Aguayo & Kaminski, 2012), but not others (Kaminski et al., 2014). In Aguayo and Kaminski (2012), phonological awareness subtest had significant correlations with all measures except TOPEL Print Knowledge (.24-.69), comprehension subtest correlations had the same pattern, with a range of .31-.69. Vocabulary subtest correlations ranged from .35 to .72.</p> <p>Predictive validity: PELI composite scores at the beginning and middle of the year had significant medium correlations (.49-.62) with all criterion measures at the end of the year (Kaminski et al., 2014). Alphabet knowledge did not correlate with all of the criterion measures. PELI is more sensitive with DIBELS Composite Score (.84) as outcome than with Core Language Score (.45), but less specific (DCS = .77, CLS = .83). Overall, DCS had a higher</p>		<p>between classrooms and individuals, finding a lot of variability at both levels of analysis. In some classrooms, children progressed by ~20 points between measurements, while in others they progressed 60-70 points. Individual growth rates varied between -20 and 120.</p>	

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
				accuracy classification (.79; vs .65 for CLS). PELI with the CLS has a higher positive predictive power (.70) than with DCS (.52), but lower negative predictive power (CLS = .70, DCS = .95). Kaminski et al., (2018) found that composite PELI score in preschool was a strong, significant predictor of DIBELS Composite Score in kindergarten, with a beta coefficient of .71.			
<p>Preschool Situational Self-Regulation Toolkit (PRISIST)</p> <p>A structured observational assessment of children's early self-regulation development, including self-regulatory behaviours around: sustaining attention; engagement; self-direction; impulsiveness; remaining within the rules of the activity; fidgeting; following social conventions such as taking turns; willingness to try even under conditions of uncertainty; emotional responses and recovery.</p> <p>Latest version: 2019</p> <p>Source: www.prsist.com.au</p>	<p>Intended cycle frequency: Not specified and not currently tested for feasibility as a progress monitoring tool during year before school.</p> <p>Online or IT-based mode: No</p> <p>Standardised scoring: Yes</p> <p>Training: Observers must complete an in-depth online training module (freely available) that provides background, administration protocols, practice ratings with feedback, and an inter-rater reliability check against a full observation and rating.</p>	<p>2. Connected with a contribute to the world Social convention, helpfulness and supportive of other children.</p> <p>4. Confident and involved learners Self-regulation including sustaining attention and resisting distraction, engagement, thoughtful and planful, self-directed, control of behavioural impulses, fidgeting, risk taking.</p>	<p>Inter-rater reliability: When completing the training module, educators must achieve a minimum threshold of consistency against a benchmark rating: mean difference in average rating ≤ 0.75 points; a correlation between item ratings of at least $r = .70$; and at least 80% of item ratings within 1 point (Howard et al., 2019).</p>	<p>Convergent validity: Concurrent association with direct assessment (HTKS, ranged from $r = .50$ to $.63$) educators' ratings of self-regulation (CBRS, r's $> .48$) and parents' ratings (CBRS, r's $< .28$). PRSIST scores were strongly related with school readiness (Brackens, r's $> .66$), (Howard et al., 2019).</p>	<p>EFA identified 2 factors for each task: Cognitive self-regulation (attention/distract, engagement, thought/plan, self-directed, helpful, risk) and Behavioural self-regulation (behaviour control, fidget/restless, social convention, emotional reaction), (Howard et al., 2019).</p>	<p>Developmental and subgroup validity: Both tasks are sensitive to age (i.e., improvements seen from 3 to 5, Howard et al., 2019) but it is not clear if this tool would be sufficiently sensitive to change within the year before school.</p>	<p>An RCT is currently underway (Howard et al., 2018, study protocol – doi:10.1186/s13063-018-2455-4). Fifty preschool centres in New South Wales, Australia, will be selected, baseline child and educator data collected, and then the participating centres will be randomly allocated to an intervention and control groups stratified by NQS rating. Primary outcomes at the child level will be two measures of self-regulation: Head-Toes-Knees-Shoulders task and the PRSIST observational assessment. Secondary outcomes at the child level will be adult-reported measures of child self-regulation, executive function and school readiness. Outcomes at the educator level will involve a survey of their perceived knowledge, attitudes and self-efficacy for supporting children's self-regulatory development.</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Age range: 3:0–5:0 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: 8 items are rated during 2 activities (memory game and curiosity box).</p> <p>Time: Memory game = 10 minutes with four children; curiosity boxes = 5 minutes/child.</p>	<p>Who administers the tool and receives feedback: Research assistants, intended for educators.</p> <p>Details included in feedback: PRSIST provides scores on discrete dimensions of self-regulation that can be used to generate actionable information linked to real-world self-regulatory behaviours and developmental sequences.</p> <p>Suitable for diverse populations: Limited testing of suitability for diverse populations. Available validity study included a small sample with normally developing and with English as first language, and no low-SES participants,</p>						
<p>Profile of Preschool Learning & Developmental Readiness (ProLADR)</p> <p>An authentic assessment of school-readiness grounded in a series of semi-structured observations, intended to complement direct child measures (IDGIs). ProLADR covers six domains: social and emotional development,</p>	<p>Intended cycle frequency: 3/year</p> <p>Online or IT-based mode: Unclear</p> <p>Standardised scoring: NA</p> <p>Training: NA</p> <p>Who administers the tool and receives feedback: Educators, families can receive feedback.</p>	<p>3. Sense of wellbeing Emotional development, self-concept, peer-related social competence and relationships, adult-related social competence and relationships. Gross and fine motor skills, physical health and well-being.</p> <p>4. Confident and involved learners</p>	NA	The publisher's website claims that the tool was developed with construct validity in mind, using prominent theoretical frameworks for child development and contributions from three child development experts. Currently no technical specifications are available on the website or from other sources.	NA	NA	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>language and literacy development, cognitive development, physical and motor development, approaches to learning, and creativity and the arts.</p> <p>Latest version: Unclear, recently published</p> <p>Source: www.myigdis.com/preschool-assessments/social-emotional-assessments/#1502390617906-f3476c97-06c0</p> <p>Age range: 3:0–5:0 years</p> <p>Cost: Request quote</p> <p>Number of items: Unclear</p> <p>Time: Ratings are based on a series of observational questions and direct prompts in a range of settings over a period of 1-month. Completion of the scale should only take 25-30 min/child.</p>	<p>Details included in feedback: ProLARD items are written as behavioural and discrete units, clearly broken into subdomains, and written to reflect year-long developmental goals. This is intended to facilitate instructional planning, however there do not appear to be any links to suggested activities or proposed changes in instruction.</p> <p>Suitable for diverse populations: NA</p>	<p>Mathematical and logical thinking, scientific thinking and problem solving. Social systems understanding, curiosity, risk taking, imagination and invention, persistence, creating, responding.</p> <p>5. Effective communication Receptive language (listening and comprehension), expressive language, emergent reading (phonological awareness, alphabet knowledge), emergent writing.</p>					
<p>Teaching Strategies Gold (TS GOLD)</p> <p>Teacher rating system (authentic performance assessment) child observation tool designed to measure the ongoing development and learning progress of children birth through kindergarten.</p>	<p>Intended cycle frequency: Collected documentation evidence (e.g., observations, artefacts, video recordings, portfolios) is summarised at three checkpoints throughout the year (fall, winter, spring).</p> <p>Online or IT-based mode: Yes</p>	<p>3. Sense of wellbeing Building relationships with others; interacting appropriately in social situations; gross-motor development; fine motor strength and co-ordination.</p> <p>4. Confident and involved learners</p>	<p>Inter-rater reliability: Lambert et al., (2015) found that all of the item, person, and Cronbach's alpha reliability coefficients across all time points and scale scores were above .90 (excellent). Inter-rater agreement between master rate and teacher was high. All were above .80,</p>	<p>Convergent validity: Multiple studies report moderate positive correlations with norm-referenced achievement instruments for language, literacy, numeracy (Soderberg et al., 2013; Lambert et al., 2013), and for social functioning and learning behaviours (Teaching Strategies LLC, 2013).</p>	<p>Confirmatory factor analysis supports the existence of a 6-factor structure. Longitudinal analysis showed scalar/strict measurement invariance indicating that TS GOLD measures the intended constructs equivalently across</p>	<p>Developmental and subgroup validity: Teachers can make valid ratings of developmental progress (correlations > .67 with age; Kim et al., 2013). Instrument is sensitive to age differences and growth over time (Lambert et al., 2014)</p>	<p>Collado, 2016 (S/ T) – a collective case study with 6 preschool educators, most children in the sample had special education needs and were the focus of the study. Teachers noted an increase in challenging behaviours as the centre shifted towards more skill-based instruction and heavy data collection. They commented that data collection for TS GOLD is a labour-intensive process (many had to fill out the forms outside of working hours at home; organised classroom activities to collect specific</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Latest version: 2010</p> <p>Source: https://teachingstrategies.com/solutions/assess/gold/</p> <p>Age range: 0:0–5:11 years</p> <p>Cost: Unclear from the website if restricted to US/South America. Can be used even without the associated curriculum (Creative Curriculum).</p> <p>Number of items: 51/53. Socio-emotional (9), physical (5), language (8), cognitive (10), literacy (12), mathematics (7). Each item is rated on a 9-point scale, with examples of expected behaviour provided for levels 2, 4, 6, and 8. Overlapping, color-coded bands indicate the typical age and/or grade-level ranges for each item measured.</p> <p>Time: Unspecified</p>	<p>Standardised scoring: Scale score tables by domain and age group.</p> <p>Training: Teachers are required to complete implementation training provided by Teaching Strategies and may also obtain inter-rater reliability certification.</p> <p>Who administers the tool and receives feedback: Educators.</p> <p>Details included in feedback: Online tool suggests activities and/or changes in instructions if aligned with Teaching Strategies Curriculum (The Creative Curriculum).</p> <p>Suitable for diverse populations: The measure is intended to be inclusive of ELLs and children with disabilities as well as typically developing children and those who demonstrate competencies beyond developmental expectations.</p>	<p>Approaches to learning (attention, curiosity, initiative, flexibility, problem solving); memory; classification skills; use of symbols to represent objects, events or persons not present; number concepts and operations; spatial relationships and shapes; measurement and comparison; pattern knowledge.</p> <p>5. Effective communication Understanding and using language to express thoughts and needs; phonological awareness; alphabet, print and book knowledge; comprehension; emergent writing skills.</p>	<p>and all but one were above .90.</p>	<p>Kim et al., (2013) reported evidence in support of unidimensionality of each domain. These findings suggest the tool is functioning as expected.</p> <p>Divergent validity: Significant associations are found across domains (i.e., TS GOLD literacy was highly associated with a direct assessment of math skills) suggesting questionable discriminant validity (Lambert et al., 2013).</p>	<p>time (Lambert, Kim & Burts, 2015).</p>	<p>With the exception of 2 language and literacy items, all items function equally well regardless of subgroup membership (e.g., disabilities, English not first language) in a sample of > 50,000 children (Kim et al., 2013).</p>	<p>observations) and became 'mechanical and meaningless' over time, leaving little time for review and reflection. While the teachers valued the descriptive information provided by recorded anecdotes, they did not find the tool helpful for individuating instruction as they typically adjusted their approach internally on a daily basis depending on the student's current level of understanding. Teachers did not find the tool useful for revealing all skills and understanding of a student and were better able to gauge the student's abilities through repeated prompting, modelling, and questioning. Many devised their own assessment tools (e.g. classroom checklists with space for anecdotes), but these also were not systematically reviewed.</p> <p>Kim, 2016 (T) – interviews with 6 Head Start teachers and 5 supervisors. Teachers found the assessment time-consuming as each had to fill out 64-66 items for 34 children. Teachers used free play time for observing, and often referred to the TS GOLD booklet to check that they locate or memorise specific indicators and associated examples of behaviour. Teachers organised activities that could elicit needed observations for TS GOLD, and interrupted free play to conduct direct assessment on individual children.</p> <p>Kim, 2018 (T) – interviews with 6 teachers and co-teachers at Head Start. Entering data daily as recommended by TS GOLD is impractical as educators were required to take time out of their breaks or take the work home when a deadline was approaching. TS GOLD</p>

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
							<p>allowed teachers to showcase their work and the progress of their students. The system is monitored by their supervisors who can see who is logging how much data. <i>Note:</i> appears to be the same study as Kim (2016).</p> <p>Little et al., 2019 (T) – interviews with 22 school district administrators and 13 state officials, and a survey for 59 teachers in North Carolina. It appears that teachers use TS GOLD with fidelity, but there are mixed reports of engagement with the tool. Only one county gathered all data together to make county-level decisions and generally the interviewees did not focus on the decision-making aspect. Only 53% of teachers reported receiving PD specific to assessment in the last 12 months, and only 47% participated in PD about using data from student assessment (in the last 12m). 92% of respondents reported using assessment results to help them individualize instruction, and 80% of respondents reported using assessment results to identify individual learning needs. Authors note that some teachers may not consider TS GOLD to be an assessment so responses may be affected.</p>
<p>Write Start! Writing Assessment</p> <p>The tool provides descriptions of early writing, including specific information about the visual, graphophonic, and semantic features.</p>	<p>Intended cycle frequency: Data analysed from 2 time points in this study.</p> <p>Online or IT-based mode: No</p> <p>Standardised scoring: Authors provide scoring categories.</p>	<p>5. Effective communication</p> <p>Provides a progression of writing form, directionality, intentionality and message content (including complexity, coherence, topic, and register).</p>	<p>Inter-rater reliability: In all models, classroom-level variances were near zero and non-significant indicating that variability in individual scores was not associated with between-classroom</p>	NA	NA	<p>Developmental and subgroup validity: Rowe & Wilson (2015) tracked changes in various writing features measured by the tool (writing form, directionality, task-message match) over</p>	NA

Tool			Psychometrics				Evidence
Description	Features	EYLF Outcomes	Reliability	Validity	Factor Structure	Social, developmental and subgroup validity	Student (S) and teacher (T) outcomes
<p>Latest version: 2015</p> <p>Source: Rowe & Wilson (2015)</p> <p>Age range: 2:6–5:11 years</p> <p>Cost: Not a commercial tool</p> <p>Number of items: NA</p> <p>Time: Unspecified</p>	<p>Training: No</p> <p>Who administers the tool and receives feedback: Research assistants, intended for educators.</p> <p>Details included in feedback: A score for each of the 4 domains of early writing covered by the assessment (writing form, directionality, intentionality, and task/message match).</p> <p>Suitable for diverse populations: Tested in a sample of low-income students, predominantly African American (98%)</p>		<p>differences (Rowe & Wilson, 2015).</p>			<p>3 years. Conventional writing forms were found to be more common in older children, but there was a lot of variability in writing forms in each age group. Over time, children's skills matured in all 4 domains ($p < .05$): word form ($b = .14$), directionality ($b = .08$), intentionality ($b = .05$), and task/message match ($b = 1.03$). The authors describe a number of patterns in various features of children's writing depending on age group.</p>	

APPENDIX D: PROFESSIONAL LEARNING SEARCH RESULTS

Organisation	Website	Times	Cost	NESA accredited	Title/s	Delivery mode	Dates	Notes
Early Life Foundations (Kathy Walker)	https://www.earlylife.com.au/	3 hrs	\$220 pp	?	Want your life back? Effective & time efficient documentation	F2F	Varied	Template included. Can also customise training.
Hydon Consultancy	https://www.hydonconsulting.com/	?	?	?	Program planning	F2F	Varied	Customised training.
KU Children's Services	https://www.ku.com.au/professional-development	6 hrs	\$190 pp	Y	Documentation: Find the Story Worth Telling	F2F	Varied	This is 'in-house training'
		3 x 1/2 sessions	100 pp	N	Pedagogy and Practice - 3 part series	Online	Varied	This is 'in-house' training
Child Australia WA	https://childaustralia.org.au/	1 hr	\$19 pp	?	STOP Drowning in papaerwork	Webinar	Varied	Can customise training
	http://www.echr.edu.au/docs/default-source/resources/ips/p/effective-curriculum-planning-and-documentation-methods-in-education-and-care-services.pdf?sfvrsn=8	N/A	Free resource	N	Effective Curriculum Planning and Documentation Methods in Education and Care Services	Publication 2012	N/A	44 page includes photos, examples, questions, etc.
		self paced	\$59 per module	?	Programming in Practice	Online - self paced	Open	
Early Childhood Australia (ECA)	http://www.earlychildhoodaustralia.org.au/	self paced	\$38.50 per module	Y	Documenting and Assessing Children's Learning 2 x parts	Online	Open	Can customise training
		?	\$25 pp	?	The Planning Cycle Rediscovered	Webinar	Open	
		?	\$25 pp	?	Keynote – Documentation as a transformative tool for change and for the becoming of democracy	Webinar	Open	

		?	\$25 pp	?	Making Curriculum Decisions: Exploring the Planning Process	Webinar	Open	
	https://www.ecrh.edu.au/resources/detail/index/talking-about-practice-planning-and-documentation	N/A	Free resource	N	Talking about practice: Planning and documentation	3 x videos	N/A	Three-part video featuring Heather Barnes talking to three educators about how they implement all the steps of the planning cycle
		?	\$140 for 4 modules	Y	Participatory Planning— The Floorbooks Approach— Claire Warden*	Online modules	Open	
Ectarc Early Childhood Training and Professional Development	http://www.ectarc.com.au/	2 hrs	\$60	Y	Purposeful program, practice, assessment and planning	Webinar	Open	
Early Childhood Resource Hub	https://www.ecrh.edu.au/home	self paced	free	?	A wide range of topics	Videos, self-learning packages, fact sheets, etc.	Open	Access to a wide range of resources, PD, etc.
Queensland Curriculum and Assessment Authority	https://www.qcaa.qld.edu.au/kindergarten	self paced	free	?	Monitoring, documenting and assessing: Kindergarten professional topic	Online package	Open	A range of resources, samples, templates, etc.
Community Child Care Victoria 2011	https://www.ecrh.edu.au/docs/default-source/resources/ips/p/child-centred-curriculum-planning.pdf?sfvrsn=4	self paced	free	N	Child-Centred Curriculum Planning (0-5 years)	Online package	Open	Training package includes questions
First Door Early Childhood Professional Learning (Queensland)	www.firstdoor.com.au	5 hrs	\$145 pp (+member ship fee)	N	Using meaningful documentation to develop the whole child	F2F (?)	Set dates	One part of a 3 part series
Australian Child Care	https://www.accco.com.au/	2 hrs	\$70 pp	?	Documentation in Early Childhood Settings	F2F	Set dates	Brisbane & Adelaide campuses

Career Options		2 hrs	\$70 pp	?	Curriculum Planning	F2F	Set dates	
Community Early Learning Australia (CELA)	https://www.cela.org.au/wp-content/uploads/2019/11/cela-pd-calendar-NOV19-JUL20.pdf		\$250pp	Y	Critical Reflection and Assessing Children's Learning	F2F	Fri, 01st Nov 2019	Reflective processes of assessment and planning - workshop. Unpack regulatory requirements for documentation.
Gowrie NSW Education Hub	https://www.gowriensw.com.au/educationhub/workshops/assessment-ready-steady-go-19/	6.5 hrs	?	Y	Programming, documentation and the assessment and planning cycle			Exploring how to document children's learning in thoughtful and meaningful ways Investigating each component of the Assessment and Planning Cycle (Element 1.3.1 of the NQS) Considering the 'Exceeding Guidance' for Quality Area One
Semann & Slattery	https://semannslattery.com/courses/1306-pedagogical-documentation-as-an-alternative-language-of-assessment-and-evaluation/	2.5	?	N	Assessment	F2F	TBA	Understand the role of pedagogical documentation A means of celebrating children's theorising, thinking and learning to an increased focus on documentation as a language of assessment and evaluation. In this provocative presentation we will explore an alternative language of documentation and liberate the ways in which we can celebrate children through this process.
Community Child Care Victoria (2011)	http://www.echr.edu.au/docs/default-source/resources/ips/p/child-centred-	N/A	Free resource	N	Child-centred curriculum planning (0-5 years); Self-guided learning package.	Free publication	N/A	Information and strategies about a child-centred curriculum.

	curriculum-planning.pdf?sfvrsn=4							
Children's Service Central (2012)	http://www.ecrh.edu.au/docs/default-source/resources/ips/p/provocations-on-assessment-in-early-childhood-education.pdf?sfvrsn=6b	Self paced	Free resource	N	Provocations on assessment in early childhood education	20 page publication 2012	N/A	Explains the term 'assessment' and why it is a pedagogical principle in the EYLF and outlines the difference between approaches to assessment.
	http://www.ecrh.edu.au/docs/default-source/resources/ips/p/whats-pedagogy-anyway-using-pedagogical-documentation-to-engage-with-the-early-years-learning-framework.pdf?sfvrsn=8	N/A	Free resource	N	What's pedagogy anyway? Using pedagogical documentation to engage with the Early Years Learning Framework	28 page publication	N/A	Looks at documentation - includes a description of pedagogical documentation and reasons why educators document.
Victorian Curriculum and Assessment Authority	http://www.vcaa.vic.edu.au/Documents/earlyyears/EYLiteratureReview.pdf	N/A	Free resource	N	Victorian Early Years Learning and Development Framework		N/A	Literature review documents the research that underpins and defines wellbeing for children from birth to five years.
Early Childhood Australia	https://www.ecrh.edu.au/docs/default-source/resources/nqs-plp-e-newsletters/nqs-plp-e-newsletter-no-73-2014-assessing-childrens-learning-work-in-progress!-(part-1).pdf?sfvrsn=6	N/A	Free resource	N	Assessing children's learning—work in progress! (Part 1)	2014 4 page e-newsletter x 2	N/A	A two-part e-Newsletter summarises important basic information about assessing children's learning and illustrates it with the perspectives and experiences of one service.

Aussie Child Care Network	https://aussiechildcarenetwork.com.au/articles/childcare-programming	N/A	online 'tutorial'	N	How To Write A Learning Story	Advertised as a tutorial	N/A	Provides a simple and easy step by step approach on what you need to include in order to complete a learning story.
Gowrie SA 2017	https://gowriesa.org.au/docs/Pedagogical-Documentation-08-02-2018.pdf	N/A	Free resource	N	Pedagogical Documentation: A South Australian Perspective	56 page resource	N/A	This e-book provide sites with a resource to support teams in understanding and engaging in pedagogical documentation to make children's and educators' learning visible.
Queensland Government	http://www.foundationsforsuccess.qld.edu.au/introduction	N/A	Free access	N	Foundations for Success	Website - examples, templates, videos, etc.	N/A	A resource to help deliver a quality early learning program for Aboriginal and Torres Strait Islander children. Provides additional guidance to the EYLF and is packed with information, inspiration & practice advice.



APPENDIX E: ADDITIONAL RATINGS FOR EACH TYPE OF ASSESSMENT TOOL

		Ease of use	Sharing with families	Sharing with other educators	Suitability for children with a range of abilities	Suitability for children who do not have English as their home language	Informing your writing of a Transition to School statement	The time needed to complete the assessment	Assessing children's learning in relation to the EYLF Learning Outcomes	Cost	Suitability for involving children in the assessment process	Usefulness for providing ongoing feedback, reflection and planning for individual children
Digital app/s	mean	4.18	4.24	4.23	4.43	3.75	3.59	3.77	4.00	3.70	3.24	4.27
	n	22	21	22	21	16	17	22	22	20	21	22
Written observations (own words)	mean	4.37	3.79	4.53	4.47	4.13	4.27	4.05	4.37	4.53	3.68	4.63
	n	19	19	19	19	16	15	19	19	19	19	19
Written observations (plus photos)	mean	4.68	4.72	4.76	4.76	4.65	4.58	3.76	4.76	4.28	4.24	4.80
	n	25	25	25	25	20	19	25	25	25	25	25
Photography only	mean	4.89	4.78	4.56	4.89	4.88	4.88	2.44	4.89	4.11	4.22	3.56
	n	9	9	9	9	8	9	9	9	9	9	9
Learning Story	mean	4.22	4.72	4.89	4.78	4.44	4.67	2.83	4.72	4.17	3.94	4.72
	n	18	18	18	18	16	18	18	18	18	18	18
Floor books	mean	4.43	4.29	4.57	4.43	4.42	3.13	3.50	3.64	4.43	4.57	3.64
	n	14	14	14	14	12	8	14	14	14	14	14
Developmental checklist	mean	4.61	3.94	4.56	4.33	3.31	4.33	3.39	3.89	4.28	3.56	4.50
	n	18	18	18	18	13	15	18	18	18	18	18
On-line assessment	mean	4.33	4.0	4.0	4.5	3.33	3.83	3.17	3.50	3.83	3.17	4.67
	n	6	6	6	6	6	6	6	6	6	6	6
Own design tool	mean	4.77	4.62	4.85	4.62	3.69	4.25	3.69	4.62	4.38	4.00	4.77
	n	13	13	13	13	13	12	13	13	13	13	13
External assessment tool	mean	5.0	5.0	5.0	4.67	3.20	4.33	4.17	4.33	3.67	4.67	5.0
	n	6	6	6	6	5	6	6	6	6	6	6
Child completed	mean	4.8	4.9	4.9	4.5	4.3	4.5	4.0	4.4	4.2	4.9	4.7
	n	10	10	10	10	10	8	10	10	10	10	10
Family completed	mean	4.38	4.84	4.62	4.62	3.54	3.31	3.92	3.38	4.08	4.15	4.31
	n	13	13	13	13	13	13	13	13	13	13	13