

# Measures of characteristics and skills associated with learning through play

## Findings from the literature

Rachel Parker, Lisa van Beeck, and Janine Callanan

Australian Council for Education Research (ACER) for  
THE LEGO FOUNDATION

*11 June 2019*

## Contents

<b>1. Introduction.....</b>	<b>2</b>
<b>2. Method .....</b>	<b>2</b>
<b>3. Key findings .....</b>	<b>6</b>
A. Conceptual framing .....	6
B. Design features.....	9
<b>4. Conclusion .....</b>	<b>11</b>
<b>5. Glossary.....</b>	<b>13</b>
<b>6. References.....</b>	<b>15</b>

## Tables

Table 1: Relevant Tools from Stage One Literature Search.....	2
Table 2: Relevant Tools from Stage Two Literature Search .....	4

## 1. Introduction

This literature review was commissioned by the LEGO Foundation to inform the revision of the Learning through Play (LtP) Experience Tool, which was developed for use at LEGO House to observe and measure the quality of children’s play experiences. This review aims to determine the extent to which the LEGO Foundation’s tool reflects good practice regarding instrument design and administration, and whether tools that measure similar characteristics using similar methods already exist.

The literature review specifications were as follows:

- Collate and synthesise literature about tools that employ similar methodologies to measure comparable characteristics to the LtP Experience Tool.
- Consider the learner age span of 5-9 years.
- Conduct searches using the five characteristics of LtP as criteria, that is, meaningful, actively engaging, socially interactive, iterative and joyful (Zosh et al., 2018).
- Include a range of tools including those used in diverse geographic contexts.
- Identify the users of the tools (researcher or practitioner) and the skills and training required to use them.
- Propose inclusions and revisions to the LtP Experience Tool based on findings from the literature.

## 2. Method

### Literature Search Stage One

The literature search was undertaken in two stages. In stage one, we identified a range of tools and independent studies of those tools that aligned with the literature review specifications. This search yielded 11 play assessment tools and rubrics, as presented in Table 1, sorted in order of relevance to the LEGO Foundation’s LtP Experience Tool. The total aggregated age range for assessment for the tools summarised in Table 1 was birth to 10 years. Most of these assessments *used* play – a play experience – to generate information about a child ( $n=9$ ). That is, they subjected a child to a play experience, generally facilitated by an adult, or encouraged a child to play, in order to prompt the child to demonstrate behaviours from which they could infer certain things about the child’s development and skills.

Table 1: Relevant Tools from Stage One Literature Search

#	Name	Measures	Purpose	Age (years)	Unit	Relevance
1	The Play Observation Scale (POS)	Cognitive and social play behaviours	Social and cognitive play profile for intervention	2–6	Groups and individual child	Defines and describes play/non-play behaviours
2	Transdisciplinary play-based assessment, 2nd Ed (TPBA2)	Cognitive, language, social, and motor skills	Skill profile for intervention	0–6	Individual child	Range of tools and informants
3	Penn Interactive Peer Play Scale (PIPPS)	Social play behaviours	Social play profile for intervention	3–5	Individual child	Lists negative and non-play behaviours
4	Social Play Continuum	Social play behaviours	Social and co-operability	4–5	Individual child	Relationship between social play and adult intervention
5	Digital Play Framework	Cognitive play behaviours	How children learn to use tech through play	4–5	Classroom	Emergent innovative research (Non-standardised)
6	ChiPPA (Child Initiated Pretend Play Assessment)	Cognitive play behaviours	Play profile for intervention	3–7	Individual child	Object substitution Cross-country validity

#	Name	Measures	Purpose	Age (years)	Unit	Relevance
7	Play in Early Childhood Evaluation System (PIECES)	Cognitive play behaviours	Cognitive development profile for intervention	0–5	Individual child	Non-facilitated (no thresholds)
8	The Affect In Play Scale (APS-P)	Affective play behaviours	Socio-emotional profile for play	6–10	Individual child	Affective dimension Age 6-10
9	Play Tools for Learning	Social play behaviours	Play profile for intervention	2–3	Individual child	No clear content relevance
10	Test of Pretend Play (ToPP)	Cognitive play behaviours	Profile for intervention	1–8	Individual child	No clear content relevance
11	Vineland Social-Emotional Early Childhood Scales (SEEC)	Social and emotional play behaviours	Socio-emotional profile	0–5.11	Individual child (parent informant)	No observation No clear content relevance

### Stage One Summary

The initial search revealed a number of tools that were relevant to the LtP Experience Tool. The first five tools listed above were deemed of high value as they included:

- descriptions of non-play behaviours
- a range of tools in addition to an observation rubric
- described the role of the facilitator
- explored new areas such as play with digital technology.

These areas were deemed valuable, as they pertained to questions raised during the review of the LtP Experience Tool, such as how to categorise non-play behaviours, and what associated tools should be developed alongside the LtP Experience Tool, in order to fully understand the other significant factors that contribute to the quality of a child’s play experience.

Table 1 includes one tool; the Digital Play Framework (Bird & Edwards, 2015), which involved observing unstructured play with digital devices. By observing children at play, they determined that there are two main objectives that drive play behaviours when children learn using digital technologies. These are:

1. Epistemic – what does this device do?
2. Ludic – what can I do with this device?

Other tools identified in the search were multidimensional, that is, they included the measurement of skills across a number of skills domains (e.g. cognitive, social, and emotional). Some used a range of instruments for data collection, such as the Transdisciplinary Play-Based Assessment 2 (TPBA-2) (Linder & Linas, 2009), which includes parent and teacher questionnaires correlated with observation rubric data to ensure greater validity and reliability of the results. The TPBA-2 is relevant to the LtP Experience Tool, as it covers a number of skills domains. The Play Observation Scale (POS) (Rubin, 2001) and the Penn Interactive Peer Play Scale (PIPPS) (Fantuzzo & Hampton, 2000) includes descriptions of non-play or negative behaviours that are relevant to the revision of the LtP Experience Tool, as these aspects are not currently included.

### Literature Search Stage Two

The stage two search involved expanding the review parameters to include tools used to measure skills and attributes such as creativity, learner agency, and learner engagement, which align with LEGO Foundation’s redefinitions of play and learning (Zosh et al. 2018). In addition, it was determined that it was important to also review instruments that are used for peer observation, as this became an

emerging potential use for the revised rubric. Finally, to acknowledge the influence of the learning environment on play experiences, we included reviews of instruments that measure classroom climate and incorporate the physical and psycho-social dimensions. This search yielded 12 tools and a further 12 were suggested for review by the LEGO Foundation. A total of 16 studies were included as relevant. Results of search stage two can be found in Table 2, organised by measurement domain.

Table 2: Relevant Tools from Stage Two Literature Search

#	Name	Measurement Domain	Purpose	Age (years)	Unit	Relevance
1	Analysing Children's Creative Thinking (ACCT) Framework	Creativity	Relationship between creativity and facilitator intervention	3–4	Children (individual, groups, and pair)	Methodology – video recorded play experiences analysed against framework
2	CRL's Five Dimensional Model of Creativity	Creativity	To create a student-level creativity profile	11–16	Groups and individual students	Student self-assessment and creativity wheel are relevant
3	Harris Whole School Creativity Audit	Creativity	Secondary school level rating for interdisciplinary creativity	N/A	School	Identifies school environment enablers for creativity
4	Student Perceptions of Control Questionnaire	Engagement	Develop construct of engagement through teacher, student and classroom observation	7–12	Student	Describes engagement, disaffection behaviours Defines engagement Age range aligns
5	Survey Measure of Math and Science Engagement	Engagement	Develop measure of maths and science engagement that reflects multidimensional engagement construct	11–19	Teacher and student	Includes notion of social engagement Includes behavioural indicators of engagement and disengagement
6	The Classroom Engagement Inventory (CEI)	Engagement	Relationship between cognitive and behavioural measures of engagement	9–17	Classroom	Distinguishes between effort and compliance regarding behavioural engagement
7	ICF-CY Framework and Child Engagement Questionnaire (CEQ)	Engagement	Correlate the ICF-CY with Child Engagement Questionnaire	3–5	Conceptual	No items provided No empirical data collection
8	inCLASS	Engagement	Establish validity of inCLASS across different demographic groups	4	Student	Training and administration process
9	Agentic Engagement Scale (AES)	Agentic engagement	Demonstrates relationship between agentic engagement and achievement and motivational support	17+	Student	Aligns with agency levels Describes behaviours associated with agency
10	The Classroom Climate Index (CCI)	Learning environment	Measure three aspects of classroom climate to validate construct	8–12	Classroom/student	Comprises student engagement, supportive teacher behaviour, and student collegial support
11	How I Feel About My School	Learning environment	Assess validity of HIFAMS questionnaire	4–8	Student	Age range aligns

#	Name	Measurement Domain	Purpose	Age (years)	Unit	Relevance
	(HIFAMS) Questionnaire					Incorporates student perception in and outside the classroom
12	SCALE: Support for Creativity in a Learning Environment	Learning environment	How learning environments support creativity	5–12	Classroom	Alignment with environments supporting LtP Can be used to audit 'student centredness'
13	Action Research Strategy for children's peer assessment skills.	Peer observation (writing skills)	Peer assessment to develop self-monitoring learners	10–12	Student	Limited relevance Associated with agency Marginal relevance to training in rubric usage
14	Instructional Practices Inventory (IPI)	Peer observation (student engagement)	Peer assessment to determine level of student engagement	5–17	Classroom	IPI used for profiling student engagement for P–12 Examples of engaged and disengaged behaviours Observation methodology
15	PATS_GO (Peer Assisted Teaching Scheme Goal Oriented)	Peer observation (effective teaching)	To align teacher peer assessments with teacher's PD goals	18+	Teacher	Addresses the 'so what /what next' after an assessment of this kind
16	The Flourishing Classroom Observation Measure (FCOM), adapted from the Classroom Observation Measure (COM).	Flourishing (positive emotion, engagement, meaning, positive relationships and accomplishment)	Assess diverse classrooms on multiple dimensions related to student well-being and academic achievement	3–5	Classroom	Methodology – observation Alignment with five characteristics of learning through play

## Stage Two Summary

Of the 16 instruments and studies reviewed for the stage two literature search, the largest proportion ( $n=6$ ) was based on student self-report questionnaires regarding values and beliefs related to their level of engagement in the classroom, or classroom climate. The second largest proportion was observation instruments ( $n=5$ ) designed to measure creativity, engagement, flourishing classrooms, and for peer (teacher) observation to determine levels of student engagement and effective teaching. The age range of subjects spanned from students attending preschool to students at college or university (approximately 3 to 18 years). The geographic contexts included Australia, Canada, England, Korea, Singapore, Sweden, and the United States. One study (Bohlmann, Downer & Williford et al. 2019) investigated the validity of the inCLASS instrument across various demographic groups such as gender, poverty status, and ethnicity, and found that it maintained similar measurement properties across different demographics.

## Five studies of interest

### 1 *The Flourishing Classroom Observation Measure (FCOM)*

FCOM was the instrument most similar to the LtP Experience Tool. It is designed to measure five elements determined as characterising flourishing classrooms: 'a) positive emotion, b) engagement, c)

meaning, d) positive relationships, and e) accomplishment’ (Waxman, Rivera, Linn, et al., 2016, p. 4). These elements are similar to the five characteristics of learning through play. FCOM is a walk-through instrument, administered over 20-minute observation periods, ‘designed to obtain multiple snapshots of classroom practices in order to provide a rich data picture’ (Waxman et al., 2016, p. 4). While the instrument is based on student wellbeing and flourishing theory (Seligman, 2011, in Waxman et al. 2016), scores are based on the frequency of positive interactions between students and teacher, activity, peers, and tasks. Unlike the LtP Experience Tool, it does not focus on how the student experiences the classroom or activity.

## 2 *Analysing Children’s Creative Thinking (ACCT)*

Robson and Rowe’s (2012) framework for ACCT includes many similar features to the LtP Experience Tool. It includes operational definitions and examples of behaviours associated with exploration, involvement and enjoyment, and persistence, which are closely linked to the characteristics meaning, engagement, and joy.

## 3 *Survey Measure of Math and Science Engagement*

The Survey Measure of Math and Science Engagement (Fredricks, Wang, Schall Linn et al. 2016) conceptualises student engagement as including four dimensions: behavioural, emotional, cognitive, and social. These dimensions align with the five characteristics of learning through play, and the survey items are a useful starting point to design supplementary instruments to support the LtP Experience Tool observation rubric (noting the differences in subject matter).

## 4 *Support for Creativity in a Learning Environment (SCALE)*

The SCALE tool (Richardson & Mishra, 2018) describes features of learning environments that foster creativity and are similar to those that are identifiable in quality learning through play environments. These features are potentially a useful foundation for developing a checklist to support the observation rubric in collecting data about important characteristics of a learning environment that supports learning through play pedagogies.

## 5 *Agentic Engagement Scale (AES)*

Reeve (2013) introduced ‘agentic engagement’ as a new educational construct, and said that ‘Conceptually, agentic engagement is a uniquely proactive and transactional type of engagement. Proactively, agentially engaged students take action before the learning activity begins’ (p. 581, 2013). By adding agentic engagement, Reeve (2013) has extended the multidimensional construct design of engagement. This research provides a reference point to the LtP Experience Tool’s descriptions of agency, which cut across all characteristics and levels. It would be prudent for the review of the LtP Experience Tool to consider how to identify both child behaviours and features of learning environments associated with a lack or an abundance of agency or choice in learning.

# 3. Key findings

## A. Conceptual framing

This section summarises the key features of studies reviewed in the literature search stage two regarding how the overall construct is conceived, its theoretical underpinnings, purpose, and importance, relationship to other concepts, and the behavioural indicators described as relating to the construct.

### *Cognitive and social play behaviours*

Key features of studies on the measurement of cognitive and social play behaviours were as follows:

- Studies were premised on theories regarding the **developmentally sequential stages of social play** in the early years (Parten, 1932 as cited in Rubin, 2001)
- Children’s play behaviours can be categorised into developmentally sequential stages regarding **cognitive play** (Piaget, 1962 as cited in Rubin, 2001)

- A child's ability to initiate and perform **pretend play** is indicative of their 'inner life, and cognitive and social abilities' (Brooke, 2004; Russ, 2005; Stagnitti & Unsworth, 2000, cited in McAloney & Stagnitti, 2009).
- Play assessments were used to diagnose a play deficit and propose an appropriate play intervention.

### *Creativity*

Key features of studies on the measurement of creativity were as follows:

- Studies were predicated on the need to **foster students' creativity and innovation to meet future workforce needs**, and to address the slow pace of change in incorporating creative skills development in schools (de Bruin & Harris, 2017).
- **Creativity is multidimensional** and involves a range of traits or habits of mind (Lucas, 2016). Lucas developed a model with five core creative habitual dispositions, namely: inquisitive, imaginative, persistent, collaborative, and disciplined (Lucas, 2016, pp. 281–282). These clearly intersect with learning through play characteristics.
- High-quality creative learning environments appear to share common features with high-quality environments for learning through play.
- One study describes fostering creativity as a **whole of school endeavour** (de Bruin & Harris, 2017) that considers 'school policies, teacher pedagogies, the nurturing of student and teacher practices and processes for creativity, school environments and local/global creative partnerships' (p. 244).
- **Creativity is linked to integrated pedagogies** as teachers practise and foster creativity through making cross-domain or interdisciplinary connections (de Bruin & Harris, 2017), for example, connecting history and mathematics or designing a unit around a meaningful question.

### *Engagement*

Key features of studies on the measurement of engagement were as follows:

- **Engagement is multidimensional**, and most conceptualisations of engagement traverse the learning through play characteristics. Hunzicker and Lukowiak (2012) describe how engaging activities are likely to be 'interactive, challenging, and relevant'. These descriptors align with the LEGO Foundation characteristics of iteration, joy, social interaction, and meaning. Bohlmann et al. (2019) established that the four dimensions that influence engagement as valid across different contexts, as positive engagement with teachers, positive engagement with peers, positive engagement with tasks, and classroom environment (related to behavioural control, teacher conflict, and peer conflict).
- **Studies describe behaviours associated with both engagement and disaffection.** Skinner, Kindermann and Furrer (2009) describe engagement as '... the quality of a student's connection or involvement with the endeavor of schooling and hence with the people, activities, goals, values, and place that compose it' (p. 494). This study includes the 'Engagement Versus Disaffection with Learning: Student Report' (Kindermann & Furrer, 2009, p. 494) in which items characterise behavioural and emotional disaffection as feeling bored, worried, nervous, discouraged, feeling bad, mad, being bothered, and frustrated. This is relevant to the LtP Experience Tool as distinguishing between engaged, disengaged or disaffected behaviour is critical to producing valid and reliable judgements from observations regarding the characteristic 'actively engaging'. Fredericks, Wang, and Schall Linn et al. (2016) characterise engagement as having four dimensions: behavioural, emotional, cognitive, and social.
- **Disaffection or disengagement behaviours are included in scales.** Valentine (2007) as cited in Hunzicker and Lukowiak, 2012, p. 103, describes six levels of engagement in the Instructional Practices Inventory. These are: 1) complete disengagement; 2) student work with teacher not engaged; 3) student work with teacher engaged; 4) teacher-led instruction; 5) student learning conversations; and 6) student active engaged learning.
- **Engagement and agency are interrelated.** Behavioural indicators of agency are also signs of engagement. Student engagement in learning is also defined as 'the extent of students' involvement

and active participation in learning activities' (Cole & Chan, 1994, as cited in Hunzicker & Lukowiak, 2012, p. 259).

### Agency

Key features of studies on the measurement of agency were as follows.

- **Conceptually, agency and engagement are closely connected.** Reeve (2006) asserted that students' inner motivational resources can either be supported or frustrated by the classroom climate, and when teachers recognise and nurture students' inner motivations, they adopt an 'autonomy-supportive motivating style' (p. 225). Reeve (2006) described autonomy-supportive behaviours as when teachers '1) listen carefully; 2) create opportunities for students to work in their own way; 3) provide opportunities for students to talk; 4) arrange learning materials and seating patterns so students manipulate objects and conversations rather than passively watch and listen; 5) encourage effort and persistence; 6) praise signs of improvement and mastery; 7) offer progress-enabling hints when students seem stuck; 8) are responsive to students' questions and comments; and 9) communicate a clear acknowledgment of students' perspectives' (p. 231). Awareness of these features is important in order to observe teacher actions that promote student agency.
- Robson and Rowe (2012), in their study of children's creative thinking, also found a positive association between agency and creativity. **Child-initiated activities were found more likely to be associated with higher engagement**, trying out and analysing new ideas, flexibility and originality, imagination and hypothesising, risk taking, and general overall persistence, than adult led activities (Robson & Rowe, 2012).
- **Consider how a lack of agency is characterised in learner behaviour.** Reeve (2009) provided examples of Controlling Motivating Style indicators, and tabulates the rationale for teachers to adopt this style, based on pressure from above, below, and within. The LtP Experience Tool does not specifically include behaviours associated with agency or a lack thereof.

### Learning Environment

Key features of studies on the measurement of classroom climate and learning environment were as follows.

- Tools to measure classroom learning environment are based on the premise of the quality and effectiveness of **student-centred learning** (Kearney, Smith & Maika, 2016; Richardson & Mishra, 2018).
- **The theory of change** is that when students feel supported, student motivation increases, and classroom performance improves (Kearney, Smith & Maika, 2016).
- **Tools include definitions of classroom climate.** 'Classroom climate has been defined as the manner in which students experience the psychological and physical characteristics of the classroom' (Gillen, Wright, & Spink, 2011, as cited in in Kearney, Smith & Maika, 2016, p. 310).
- **The learning environment construct is multi-dimensional.** Kearney, Smith and Maika (2016) describe it as including three important aspects: 'collegial support, student engagement, and supportive teacher behaviour' (p. 319). Richardson and Mishra (2018) propose that classroom environment comprises the physical environment, learning climate, and learner engagement.

### Peer observation

Key features of studies on peer observation instruments were as follows.

- **Peer-observation activities are seen as an opportunity for professional development, not evaluation.** Hunzicker and Lukowiak (2012) summarise peer observation as a way to provide teachers with ongoing professional development and opportunities to share knowledge, offer

suggestions and support, and recognise effective teaching. The limitations of the Instructional Practices Inventory (IPI) were stated as:

1. Coding is dependent on the observer's understanding of students' developmental levels and prior experiences, and may be influenced by the observer's biases.
2. Students can (and often do) appear to be actively engaged in learning when actually they are not.
3. The IPI codes describe student activity, but not student learning outcomes (Hunzicker & Lukowiak, 2012, p. 105).

These are important to note as potential limitations of using the *Learning through Play Experience Tool*, if used as a peer-observation instrument.

- **Users need to be well informed about quality benchmarks.** Boon (2015) lists the necessary enabling factors for peer assessment as:
  1. students knowing what a competent piece of written work looks like
  2. they are shown how to assess writing and given practice opportunities to do so
  3. suitable scaffolds such as prompts, are provided, so their assessments focus on relevant text-level features.

These features could be incorporated into teacher training to use the LtP Experience Tool. Teachers may be more successful in using the tool if they:

- understand what high-quality learning through play looks like
  - are shown how to assess learning through play and given practice opportunities to do so
  - are provided with prompts that direct their attention towards the key behaviours.
- **Consideration should be given to next steps after a peer assessment.** For the Peer Assisted Teaching Scheme Goal Oriented tool (PATS\_GO) (Drew, Phelan & Lindsay, et al. 2017), two or more peers work together to analyse past and present teaching evaluation data to determine development goals, and then design, execute and evaluate a strategy for achieving them. Instrumentation and associated materials should be clear on what should happen next if teachers find that their classrooms score low on the five characteristics.

## B. Design features

The measures of skills and characteristics associated with learning through play reviewed possessed a number of common design and administration features that were central to claims made about their validity and reliability. These are summarised as follows.

### *Specific definitions of constructs to be measured or observed*

High-quality assessment instruments included specific definitions of the constructs to be measured or observed. For example, the Play Observation Scale (Rubin, 2001) includes social and cognitive play and includes specific definitions of each of these terms, and the subcategories within them. It is also considered good practice to seek input and consensus on definitions from key stakeholders. Further, the roles of those participating in the observation, such as teachers, aides, guides, and facilitators, were described clearly. For example the TPBA 2 (Linder & Linas, 2009) describes distinct roles for the play facilitator, parent, and rater. Finally, the desired inferences from results of observations were clearly stated and evidence collected was sufficient to support them. The Child Initiated Pretend Play Assessment (ChiPPA) (Stagnitti, 2007) provides a strong illustration of this point. In the guidelines on how to administer ChiPPA, Stagnitti (2007) explains how behaviours have distinct play action codes, and from the frequency of scoring particular play codes (B, R and F codes), the assessor can accurately make inferences about the child's play ability.

### *Comprehensive domain coverage*

Instruments reviewed included broad and comprehensive domain coverage. All reasonably anticipated and relevant behaviours associated with the construct were described on the instrument and able to be

coded on it. Referring to the Play in Early Childhood Evaluation System (PIECES), Kelly-Vance and Ryalls (2008) said that that 'every play behavior produced by the child can be classified on the core subdomain' (p. 552). The Play Observation Scale's description of social play types are broad, yet specific enough to cover all types of play behaviours within the category (Rubin, 2001). For example,

'Solitary Play: The child plays apart from other children at a distance greater than three feet (one meter). S/he is usually playing with toys that are different from those other children are using. The child is centered on his/her own activity and pays little or no attention to any children in the area. If the child is playing in a small area the three-foot rule is often not applicable. In such cases the observer must rely upon the relative attentiveness of the child to others in his/her social milieu' (Rubin, 2001, p. 3).

Rubin's definition includes a general distinguishing feature (distance from other children), but provides an alternative feature should the distance be inapplicable.

#### *Distinction between components and levels*

Instruments reviewed included clearly separated components and/or scales. The components measured distinct aspects of behaviour or performance, and evidence collected as part of the observation aligned clearly with a single category or component of the instrument. For example, the PIPPS has three categories – play disruption, play disconnection, and play interaction – and lists of behaviours associated with each construct. Exploratory factor analysis revealed that these three components are three distinct constructs (Fantuzzo & Hampton, 2000, p. 608).

#### *Parallel language and structure across component levels*

Instruments reviewed employed comparable language and structure across the different component levels. They used consistent descriptive language across levels within a component, and included behavioural descriptors in all levels of the instrument, or none of them. The Play Observation Scale (Rubin, 2001), consistently described play types across the two main categories commencing with the central idea that the type is based upon, followed by an explanation, and an example. All of the tools reviewed included behaviours for all categories. The Social Play Continuum (Broadhead, 2001) included classifications of behaviours, such as 'L' for language, 'A' for action, and 'RA' for reciprocal action, and these classifications were used consistently across the framework.

#### *Attention to observer cognitive load*

Instruments reviewed described how the administration protocols paid attention to observer cognitive load. They acknowledged how observers must be able to process the amount of behaviour or performance without overwhelming cognitive load, which leads to biased observations, inaccuracy, overgeneralisation, and data loss. Observations were separated into shorter segments or pieces and many of the tools reviewed included specific times (15–20 minutes) and chunks of time (often 5 minutes) allocated to specific observational activities. Bohlmann et al. (2019) described the administration process as follows: 'During each classroom visit, data collectors conducted four 15-minute cycles every hour (10 minutes of observation, 5 minutes of coding) alternating between selected children until the end of the observation (approximately 4 hours), resulting in four inCLASS cycles per child' (p. 169). Waxman et al. (2016) stated that they found a 20-minute observation timeframe to be ideal for research purposes, sufficiently allowing them to capture instructional quality and classroom environment.

Approaches to managing observer cognitive load were:

- separating observations into shorter segments or pieces
- designing instruments with the minimum number of components and scales required to observe and code behaviours
- focusing instruments on the essential elements of the activity or performance with no more than the minimum number of levels needed to describe them.

### *Comprehensive observer training and clear administration protocols*

Most studies described the type of training and prior experience required for administrators to use the instrument effectively. It is essential that instruments are administered consistently, as this is the basis for valid inferences. The key features of high-quality observer training and administration protocols were:

- observers were required to achieve a predetermined accuracy standard before live coding
- observers were trained using materials and examples that are as similar to the operational context as possible
- the accuracy level of observers' instrument use was evaluated before live observation, and reassessed at appropriate intervals thereafter
- using multiple observers and multiple observation sessions of activity of the same subject increased the reliability of observation results.

Bohlmann et al. (2019) described the training process for inCLASS which aligns with the principles above, as follows.

'All data collectors were required to attend an intensive training session about the observational measure and reliably code video training clips before observing live in the field. The 2-day training involved detailed review of all content/dimensions, combined with watching, coding, and discussing five training clips. At the end of training, data collectors were required to code five reliability clips independently (without discussion), and score within one point of a master code on 80% of the dimensions to be deemed reliable and certified for live data collection. If data collectors did not meet this standard of reliability, they received individual consultation and were required to repeat reliability with new clips prior to live data collection' (p. 169).

## 4. Conclusion

The following summary of concluding statements and recommendations based on the literature review is organised thematically.

### *Uniqueness of the LtP Experience Tool*

The literature review finds that instruments exist that conceptualise and measure similar play skills and characteristics to the LtP Experience Tool. However, none of the instruments reviewed measured the **quality of student experiences** of learning through play. The LtP Experience Tool requires the user to make judgements about the child's experience of play from behaviour they demonstrate as they interact with the activity, facilitator, peers, and learning environment. It requires users to think about play experiences from a child's perspective instead of seeing a child as the subject of assessment. This is a unique feature of this tool.

Behaviours associated with the characteristics of exploration, involvement and enjoyment and other related constructs have been described in various instruments.

FCOM (Waxman, Rivera, Linn, et al., 2016) examines elements of learning environments such as positive emotion, engagement, meaning, and positive relationships, which are similar or identical to the LtP Experience Tool, suggesting that the LEGO Foundation model aligns with prior research.

### *Cultural validity*

Instruments are appropriate when they are designed and developed to consider the culture of the child who is to be evaluated (Fantuzzo & Hampton, 2000). Culturally valid instrument design involves understanding which competencies and skills are valued by the child's community; understanding community perceptions of child development; and developing categories based on these understandings (Fantuzzo & Hampton, 2000). Most of the tools reviewed in this literature review were designed for use

in Australia, Canada, the United Kingdom and the United States. One tool, ChiPPA, has been used in seven countries, including Brazil and within Australian Indigenous communities. Stagnitti (2007) stated that no children have experienced difficulty recognising the toys as toys, provided they were gender neutral and not 'fad' toys. Stagnitti (2007) also found that four year olds prefer toys that 'look like they belong together' (p. 2), and that three year olds were less particular about this aspect. Broadhead (2006) described how professional discourse about supporting children's activities were consistent between Italy and England.

We propose that the revised LtP Experience Tool is piloted in a range of culturally diverse settings to determine how the construct validity is maintained and how the behavioural indicators perform. It is possible that international adaptations for behavioural indicators may need to be included in a revised version of the LtP Experience Tool.

### User expertise

The literature generally described a level of expert knowledge required to use the tools in terms of the training provided to these users. Users (also referred to as observers or data collectors) were required to attend training programs over multiple days and achieve a specified standard of coding accuracy, or above a certain benchmark on a test related to the instrument and its properties (McAloney & Stagnitti, 2009; Waxman, Rivera, & Linn et al. 2016; Hunziker & Lukowiak, 2012). For instruments such as peer observation tools, where users collected qualitative data through guided interviews and observations, users were required to possess knowledge about each of the dimensions of the framework and how they manifest in different activities (Drew, Phelan & Lindsay, et al. 2017). Kelly-Vance and Ryalls (2008) stated that knowledge of child development and the developmental discourse of play is required for consistent use of certain tools. *The Framework for Teaching Evaluation Instrument 2013 Edition* (Danielson, 2013) describes how users 'speak [the] language' of the framework (p. 6), implying that the terminology of the tool is specific to it, and has to be understood by users to use successfully. It also implies an affective dimension; an alignment of values and attitudes in order to communicate effectively with others.

We propose that users of the revised LtP Experience Tool will need to be 'expert' insofar as they will need to:

- have experience in education, as an early childhood educator, teacher or education researcher; and
- undergo training in the administration of the tool, reaching a certain level of coding accuracy using authentic videos prior to live coding.

Ideally, users will be familiar with learning through play and pedagogies that view children as capable and competent, including knowledge of child development and familiarity with theories and discourse about the developmental stages of play. These areas can be partly covered in training.

### Additional tools

If it is deemed important to include questionnaires to gather supporting data regarding characteristics that do not appear to be observable, there are a number of relevant instruments included here that may be useful. These include:

- The Survey Measure of Math and Science Engagement (Fredricks et al. 2016), which provides a useful conceptual starting point to design supplementary instruments to support LtP Experience Tool observation rubric.
- SCALE (Richardson & Mishra, 2018), which offers a useful foundation to develop a checklist of features of the learning environment including physical, learning climate, and learning engagement, to correlate with observation rubric data.
- Instruments related to creativity, which have some relevance to learning through play, as creativity is linked to integrated or interdisciplinary pedagogies.

### Non-play or behaviours outside the current LtP Experience Tool scope

Non-play and other behaviours have been articulated in instruments such as PIPPS, POS, the Student Perceptions of Control Questionnaire, and the IPI. The AES highlights that consideration should be given to how to identify and code behaviours and regarding a lack of agency or choice in learning. This data could be gathered through administering different instruments.

### Use for peer observation

Based on insights from peer-observation tools, we suggest that teachers will be more successful at using the LtP Experience Tool if they:

- understand what high-quality learning through play looks like
- are shown how to assess learning through play and given practice opportunities to do so
- are provided with prompts that direct their attention toward the key behaviours.

### Design features of high quality measures of play skills and characteristics

High quality (valid and reliable) tools to measure play skills and characteristics included:

- specific definitions of constructs to be measured or observed
- comprehensive domain coverage
- distinction between components and levels
- parallel language and structure across component levels
- attention to observer cognitive load
- comprehensive observer training and clear administration protocols.

Evaluation and revisions to the LtP Experience Tool must consider the extent to which these design features have been included.

## 5. Glossary

Many of the terms used in this review are defined in different ways by various bodies of education research. This glossary clarifies how each term is used in this review.

<b>activity</b>	A play experience that children participate in, which runs for a particular duration that may or may not be specified.
<b>agency</b>	The capacity of a child to act independently and make choices freely.
<b>assessment</b>	The act of gathering information about a phenomenon to make an informed judgement about its value, quality or importance. Formative assessments are methods used by teachers to monitor student learning and provide feedback to teachers and students on the impact of instruction and learning achievement, and to inform modifications to teaching and learning activities. Assessments are often administered during, or in the process of learning, rather than at the completion of a program or unit. Summative assessments are administered generally at the end of a course, program, or key stage, in order to evaluate student learning normatively, or against a standard or benchmark.
<b>behavioural indicators of non-play</b>	Rubin (2001) described non-play behaviours as ‘unoccupied behavior, onlooker behavior, conversations with teacher and/or peers, transitional, aggressive, rough-and-tumble, hovering, and/or anxious behaviors’ (p. 3).

<b>classroom climate</b>	The ways in which students experience the psychological and physical characteristics of the classroom.
<b>construct</b>	Constructs cannot be measured directly and are therefore represented as a set of variables used to create indices or scales, for example socioeconomic status, enjoyment of reading, or self-efficacy in mathematics (ACER-GEM & UIS, 2017).
<b>continuum</b>	A continuous sequence of hierarchical elements that are related to each other.
<b>domain</b>	The area of learning that is the focus of an assessment. This may be a curriculum area (e.g. mathematics or science), or more generic areas of learning (e.g. reading, writing or problem-solving) (ACER-GEM & UIS, 2017, p. 55).
<b>experience (noun)</b>	An event or occurrence that leaves an impression on someone.
<b>experience (verb)</b>	To encounter; to access a feeling in response to an event.
<b>level(s) zero/minus of the tool</b>	A level of '0' or a minus symbol ' - ' is used to indicate instances when there was no opportunity to observe the characteristic in question, or the behaviours were considered as non-play: non-compliant, negative, or resistant behaviour.
<b>learning through play</b>	A pedagogy that combines playful, child-directed activity, with intentional facilitation on the part of the educator, to foster a broad range of learning outcomes.
<b>norm-referenced assessment</b>	An assessment that compares the results of the test taker to the expected performance of their peers of a similar age, experience level, or cohort.
<b>play facilitator</b>	A specialist trained by the LEGO Foundation to guide and support play activities at LEGO House.
<b>play environment</b>	A place where learning through play is deemed to occur.
<b>play experience</b>	In early childhood education, a play experience creates an impression on participants and ideally provides the conditions for learning due to the appropriate nexus between environment, facilitator, and activity.
<b>rubric</b>	A rubric is a scoring tool with described hierarchical levels, usually in form of a grid, used to evaluate the quality of child's performance or response within a given domain.
<b>standardised assessment</b>	A formal assessment that has been designed to measure a child's abilities, etc., using a tool that has been selected and checked empirically, for which norms have been established, uniform methods of administering have been developed, and which may be scored with a relatively high degree of objectivity
<b>tool</b>	An instrument used to measure abilities, skills, characteristics, or features of learning environments. The terms 'instrument' and 'tool' are used interchangeably in this review.
<b>typical behaviour/atypical behaviour</b>	Typical behaviour is behaviour demonstrated by a child in accordance with the predicted rate and sequence for the child's

age and developmental stage. Atypical behaviour is behaviour demonstrated by a child that does not align with the predicted rate and sequence for the child's age and developmental stage.

## 6. References

- Australian Council for Educational Research - Centre for Global Education Monitoring, and UNESCO Institute for Statistics (ACER-GEM). Principles of Good Practice in Learning Assessment. *Australian Council for Educational Research Centre for Global Education Monitoring and UNESCO Institute for Statistics, 2017*. Retrieved from:  
<http://uis.unesco.org/sites/default/files/documents/principles-good-practice-learning-assessments-2017-en.pdf>
- Bird, J., & Edwards, S. (2015). Children learning to use technologies through play: A Digital Play Framework. *British Journal of Educational Technology, 46*(6), 1149–1160.  
<https://doi.org/10.1111/bjet.12191>
- 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.  
<https://doi.org/10.1016/j.appdev.2018.08.007>
- Boon, S. I. (2015). The role of training in improving peer assessment skills amongst year six pupils in primary school writing: An action research enquiry. *Education 3-13, 43*(6), 666–682.  
<https://doi.org/10.1080/03004279.2013.856930>
- Broadhead, P. (2001). Investigating sociability and cooperation in four and five year olds in reception class settings. *International Journal of Early Years Education, 9*(1), 23-35.  
<https://doi.org/10.1080/713670680>
- Broadhead, P. (2006). Developing an understanding of young children's learning through play: The place of observation, interaction and reflection. *British Educational Research Journal, 32*(2), 191–207.  
<https://doi.org/10.1080/01411920600568976>
- Chessa, D., Di Riso, D., Delvecchio, E., Salcuni, S. & Lis, A. (2011). The Affect in Play Scale: Confirmatory factor analysis in elementary school children. *Psychological Reports, 109*(3), 759–774.  
<https://doi.org/10.2466/09.10.21.pr0.109.6.759-774>
- Conner, J., Kelly-Vance, L., Brigitte Ryalls & Mary Friehe (2014) A Play and Language Intervention for Two-Year-Old Children: Implications for Improving Play Skills and Language, *Journal of Research in Childhood Education, 28*:2, 221-237, DOI: 10.1080/02568543.2014.883452
- Danielson, C. (2013). *The framework for teaching evaluation instrument* (2013 ed.). Princeton, NJ: The Danielson Group.
- de Bruin, L., & Harris, A. M. (2017). Developing creative ecologies in schools: Assessing creativity in schools. *Australian Art Education, 38*(2), 244–261. Retrieved from  
<https://www.arteducation.org.au>
- Drew, S., Phelan, L., Lindsay, K., Carbone, A., Ross, B., Wood, K., ... Cottman, C. (2017). Formative observation of teaching: Focusing peer assistance on teachers' developmental goals. *Assessment and Evaluation in Higher Education, 42*(6), 914–929.  
<https://doi.org/10.1080/02602938.2016.1209733>

- Fantuzzo, J. W., & Hampton, V. R. (2000). Penn interactive peer play scale: A parent and teacher rating system for young children. In K. Gitlin-Weiner, A. Sandgrund & C. Schaefer (Eds.), *Play diagnosis and assessment* (2nd ed., pp. 599-620). Hoboken, NJ: John Wiley & Sons.
- Fredricks, J. A., Wang, M-T., Schall, J. L., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction, 43*, 5–15. <https://doi.org/10.1016/j.learninstruc.2016.01.009>
- Hunzicker, J., & Lukowiak, T. (2012). Effective teaching and student engagement in the college classroom: Using the Instructional Practices Inventory (IPI) as a tool for peer observation and self-reflection. *Journal on Excellence in College Teaching, 23*(1), 99–132. Retrieved from <http://celt.muohio.edu/ject>
- Kearney, W. S., Smith, P. A., & Maika, S. (2016). Asking students their opinions of the learning environment: An empirical analysis of elementary classroom climate. *Educational Psychology in Practice, 32*(3), 310-320. <http://dx.doi.org/10.1080/02667363.2016.1173015>
- Kelly-Vance, L., & Ryalls, B. O. (2008). Best practices in play assessment and intervention. In *Best Practices in School Psychology V* (vol. 2, pp. 549-559). Retrieved from [https://www.academia.edu/2236966/33\\_Best\\_Practices\\_in\\_Play\\_Assessment\\_and\\_Intervention](https://www.academia.edu/2236966/33_Best_Practices_in_Play_Assessment_and_Intervention)
- Linder, T., & Linas, K. (2009). A functional, holistic approach to developmental assessment through play. *Zero to Three, 30*(1), 28–33.
- Lucas, B. (2016). A five-dimensional model of creativity and its assessment in schools. *Applied Measurement in Education, 29*(4), 278–290. <https://doi.org/10.1080/08957347.2016.1209206>
- McAloney, K., & Stagnitti, K. (2009). Pretend play and social play: The concurrent validity of the Child-Initiated Pretend Play Assessment. *International Journal of Play Therapy, 18*(2), 99–113. <https://doi.org/10.1037/a0014559>
- Reeve, J. (2006). Teachers as facilitators : What autonomy-supportive teachers do and why their students benefit. *The Elementary School Journal, 106*(3), 225–237. <http://dx.doi.org/10.1086/501484>
- Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist, 44*(3), 159–175. <https://doi.org/10.1080/00461520903028990>
- Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology, 105*(3), 579–595. <https://doi.org/10.1037/a0032690>
- Richardson, C., & Mishra, P. (2018). Learning environments that support student creativity: Developing the SCALE. *Thinking Skills and Creativity, 27*, 45–54. <https://doi.org/10.1016/j.tsc.2017.11.004>
- Robson, S., & Rowe, V. (2012). Observing young children’s creative thinking: Engagement, involvement and persistence. *International Journal of Early Years Education, 20*(4), 349-364. <https://doi.org/10.1080/09669760.2012.743098>
- Rubin, K. H. (2001). *The Play Observation Scale [Revised]* [Test]. College Park, MD: University of Maryland.

- Skinner, E. A., Kindermann, T. A., & Furrer, C. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493-525. <https://doi.org/10.1177/0013164408323233>
- Stagnitti, K. (2007). *Child-Initiated Pretend Play Assessment* [Information sheet]. Retrieved from <https://www.learntoplayevents.com/wp-content/uploads/2016/03/Information-on-ChIPPA.pdf>
- Wang, Z., Bergin, C., & Bergin, D. A. (2014). Measuring engagement in fourth to twelfth grade classrooms: The Classroom Engagement Inventory. *School Psychology Quarterly, 29*(4), 517–535. <https://doi.org/10.1037/spq0000050>
- Waxman, H. C., Rivera, H. H., Linn, B., Padrón, Y. N., Rollins, K. B., Boriack, A. W., & Alford, B. A. (2016). Development of an observation instrument to measure flourishing learning environments. *Journal of Chinese Studies, 1*(1), 1–7. <https://doi.org/10.1186/s40853-016-0005-9>
- Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., ... Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology, 9*(1124), 1–12. <https://doi.org/10.3389/fpsyg.2018.01124>