Identifying Standardized Instruments for Measuring Play's Effect on Child Development
Findings from an Extension of a Systematic Review

•

Farahiyah Wan Yunus, Sakinah Idris,
Siti Noraini Asmuri, Bess Fowler, and
Muhammad Hibatullah Romli

The authors contend that children benefit from play as a form of intervention and as a means of fostering their cognitive, social, and physical growth. They review several standardized instruments developed over the last fifty years to assess this benefit of play on child development. They identify twenty-one such play measures, the majority of them applicable to children under twelve. Some of the measurements can be conducted by children, but most are proxy rated and employ some form of observation. Over half require no or minimal training. Others can be completed in less than thirty minutes. More evaluations exist for the earlier instruments, but these measurements lacked psychometric evidence to back them up. The authors note they based their selections on the distinct characteristics and features of the measures, and they offer their conclusions to assist users in choosing the best systematic instruments for their specific needs. **Key words**: assessing play, child development, evaluating play assessment instruments, meta-analysis of play measurement, reviewing measurement tools

Introduction

Pʟᴀʏ ɪs ᴀ ᴄᴏᴍᴘʟᴇx ᴀᴄᴛɪᴠɪᴛʏ. Definitions of play vary, but it can generally be conceptualized as an ever-evolving set of behaviors in which the players typically desire to participate irrespective of their perceptions of reality or their situation. Such play encompasses a range of physical, linguistic, cognitive, social, and

psychological abilities that fit the players' cultural background and surroundings (Stagnitti 2004). Play is also an important daily activity, particularly for children and adolescents. Although age progressively inhibits an element of play, its essence—its playfulness—persists throughout life. Because play is so important an activity, we need the right tools, those with acceptable psychometric properties, to help identify problems in the linguistic abilities, cognitive development, communication skills, and social aptitude and maturation of individual players (Salcuni, Mazzeschi, and Capella 2017; Bundy 2010).

Psychometric properties characterize the methodological quality of the evaluation tools. They include validity (measures what is intended to be measured), reliability (ensures the tools are stable over time), and responsiveness (determines whether the tools can detect changes in conditions). Consequently, therapists can use play as an intervention medium as long as they identify the objectives and the features they use to evaluate whether the objectives have been accomplished (Bundy 2010; Dankiw et al. 2020; Gomes, Maia, and Varga 2018; Novak and Honan 2019). The use of play as an intervention or therapy and a medium to measure outcomes—one that yields positive findings (Francis et al. 2022; Dijkstra-de Neijs et al. 2021; Kent et al. 2020) and guides or structures play—has more benefits than unstructured or nonguided play (Skene et al. 2022). However, to measure play properly, it is crucial to ensure that a meaningful outcome and an intervention plan can be devised from the play activity (Salcuni, Mazzeschi, and Capella 2017; Stagnitti 2004) and help the family members understand and meaningfully engage in the play activity (Foulds 2023). Therefore, selecting an appropriate measurement constitutes the first step in developing and implementing an effective intervention plan.

Evaluating play is one of the crucial steps in play therapy. Using standardized assessments to do so has become important for establishing credible outcomes from the findings. Using a standardized assessment is part of practicing evidence-based activity, facilitating effective communication, planning for intervention, and providing measurable outcomes (Duncan and Murray 2012; Unsworth 2011). However, in practice, using play assessment has attracted criticism as insufficient because of its limited focus—for example, only on development (Salcuni, Mazzeschi, and Capella 2017). In addition, some studies found the use of standardized play assessments to be stressful and unmotivating for the children involved (Cubas and Levratto 2019). Some standardized play assessments also suffer from limited exposure and access, as well as from a lack of understanding about how to use them (Kuhaneck et al. 2013; Cubas and Levratto 2019).

There exists only limited literature in play evaluation to inform the use of available instruments—our term for any measure, process, or procedure for gauging the effects of play—and the majority of these reviews are nonsystematic. A critical review by Brooke (2004) analyzes only three play assessments. Bulgarelli and colleagues (Bulgarelli et al. 2018) have conducted a systematic search and found twenty-nine play assessments.

Nonetheless, the listed instruments incorporate play—that is, they use play and evaluate play-related aspects such as functional play, games with rules, constructive play, dramatic or pretend play, and exploratory play. Some instruments are play based—that is, they use play but to evaluate nonplay components such as physical or locomotor, social, and cognitive skills. Other instruments involve nonplay assessment—that is, they do not use play activity and they evaluate nonplay tasks such as unoccupied behavior; onlooker behavior; active conversations with teacher and peers; transitional, aggressive, rough-and-tumble, hovering, or anxious behaviors; and talking or moving from one place to another. These reviews merely described the assessments without evaluating them critically.

Play-based assessment differs from mere play assessment, because—although play-based assessment employs play activity—it also evaluates nonplay aspects such as physical and cognitive ones. In Cubas and Levratto's (2019) review, they introduced ten assessments of the cognitive, social, and emotional aspects of play but did not conduct a systematic search for them. Conducting an unsystematic search for such assessment tools might have introduced a preference bias, which could have led researchers to miss potentially valuable assessments and evidence of application.

We found two systematic reviews of play instruments. O'Grady and Dusing (2014) investigate available play-based instruments to evaluate children's cognitive and motor skills. However, the instruments gathered are a mix of play and nonplay assessments and are restricted to physical therapy. A review by Romli and Wan Yunus (2020) identifies only play instruments used by occupational therapists. This study excludes valuable generic play instruments that play therapists might find interesting. Therefore, we need a more comprehensive review that collects all potential play instruments. Thus, we aim our review at identifying and collating empirical evidence based on predetermined criteria as it pertains to that play which serves as an intervention or otherwise impacts the cognitive, social, and physical development of children.

Materials and Methods

This article expands the systematic review by Romli and Wan Yunus (2020). We registered the protocol with the International Platform of Registered Systematic Review and Meta-Analysis Protocals or INPLASY (INPLASY202040156), and it is available in full at https://doi.org/10.37766/inplasy2020.4.0156) and the International Prospective Register of Systematic Reviews or PROSPERO (CRD42020170370). We previously described the methodology in detail in an earlier publication (see Romli and Wan Yunus 2020) and followed Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) checklists (Shamseer et al. 2015). However, we provide a concise and additional description of the original protocol methodology here as well.

Study Identification
Romli and Wan Yunus (2020) conducted the initial review. They systematically searched six databases of research studies including Academic Search Complete, CINAHL, MEDLINE, Psychology and Behavioral Science Collection, Scopus, and ASEAN Citation Index using key words generated by the authors of the review. The key words were generated based on previous literature and discussion among authors. They used key words related to play (e.g., playthings, play based, etc.), psychometric properties (e.g., reliability, validity, etc.), and evaluation (e.g., assessment, measurement, etc.) employing Boolean operators, parenthesis, truncation, and wild cards. They imposed no restriction on specific disciplines, allowing the gathering of potential play instruments from any discipline. They conducted the search on January 21, 2020. They followed with a manual search of the reference list of eligible articles and articles they possessed or knew about.

Eligibility Criteria
With slight modifications, we developed our inclusion and exclusion criteria based on Romli and Wan Yunus (2020). The already defined criteria that we maintained for this review were: studies about the instruments for leisure types of play, not including competitive play or sports; instruments generally evaluating play; studies investigating the psychometric properties of the instruments; and instruments solely applied to play that were not multidimensional. Our criteria excluded any instrument that was not a primary study (such as a review or editor's note); had no full text available; had no full text available in English; consisted of gray literature (such as a thesis, book, or conference), or was an

article appearing in a journal without peer review. We excluded occupational therapy play instruments because they had been investigated extensively previously by Romli and Wan Yunus (2020). In addition, we have excluded play instruments not intended for use with children and adolescents.

Instrument Selection

Two authors of this article independently screened the retrieved articles by title, abstract, and full text according to our criteria. We then selected articles and searched their full texts for potential play instruments. We scrutinized any instruments we identified for their status as nonoccupational, therapy-developed instruments by reviewing the original or earliest study in which they appeared. We then selected the eligible instruments for final analysis.

Data Extraction and Quality Appraisal

We extracted each instrument we included for its clinical utility and psychometric properties using an Internet search based on the instrument's name to find its original article according to its first publication.

The information on clinical utility collected includes details about the instrument's purpose, population, administrative aspects (i.e., training requirement, duration, assessor, scoring procedure), and language availability.

We used evidence of the instrument's psychometric properties for our quality appraisal rating. We appraised the quality of the play instruments we included using the quality tool from Francis and his colleagues (Francis et al. 2016) because this tool evaluates the instrument, while other tools evaluate overall studies (Rosenkoetter and Tate 2017). We developed the tool through a literature review and twenty-four cognitive interviews with clinicians and researchers before we shaped it into its final form, which consists of eighteen items. The tool focuses on six domains: conceptual model (three items); content validity (three items); reliability (two items); construct validity (four items); scoring and interpretation (three items); and respondent burden and presentation (three items). We scored items as either "0=criterion not met" or "1=criterion met." We calculated no total score because the tool has an unequal weight between its item and domain, so we interpreted and presented them individually (Francis et al. 2016). The tool has overall moderate inter-rater reliability ($k = 0.70$).

Data Synthesis

We summarize the psychometric properties for each assessment from the orig-

inal article, articles found in the systematic searching, and articles found in additional Internet searching. The summary provides evidence of the validity and reliability of the assessment. We present comparisons of the assessment's popularity, psychometric evidence pattern, and usability.

Results

We found 1,098 articles through electronic and manual searches. To extract available instruments, we evaluated 52 articles. We extracted a total of thirty-three instruments. However, eight of them are occupational therapy instruments and have been previously discussed in Romli and Wan Yunus (2020), and four are not for children or adolescents. These were excluded from this review, leaving only twenty-one instruments for analysis (see figure 1).

We included the characteristics of the instruments we present in figure 2. Most of these were developed independently, but a few (n=2) were based on previously established instruments. For example, the Children's Play Therapy Instrument (CTPI) was first developed for use on children with a clinical diagnosis who have received therapy. The later version, Children's Developmental Play Instrument (CDPI) was developed as a briefer version of CTPI for use with mainstream children. Howes Peer Play Scale is an enhanced version of Parten's Social Play Hierarchy. Howes retains the six items from the original Parten version but refines its scoring criteria with a four-point scale rather than a subjective percentage. Additional information for both instruments can be found in figure 2.

Several instruments are proxy rated (n= 18) either by the professionals such as play therapists, clinical psychologists, speech therapists, and teachers or by peers and parents based on observation, report, or perception (e.g., Affect of Play Scale, CDPI, CPTI, Children Playfulness Scale, Howes Peer Play Scale, Lunzer's Scale of Play Behavior, Mature Make Believe Play Observational Instrument, Parten's Social Play Hierarchy, Penn Interactive Peer Play Scale, Play Behavior Observation System, Play Performance Scale, Revised Class Play, Singer's Observational Play Instructions and Imaginative Play Predisposition Interview, Symbolic Play Test, The Play Checklist, The Social Play Record, The Test of Pretend Play, and Transdisciplinary Play-Based Assessment). Other instruments (n=3) are self-administered questionnaires. Several assessments (n=6) are commercially available in books or manuals (e.g., The Social Play Record, Symbolic Play
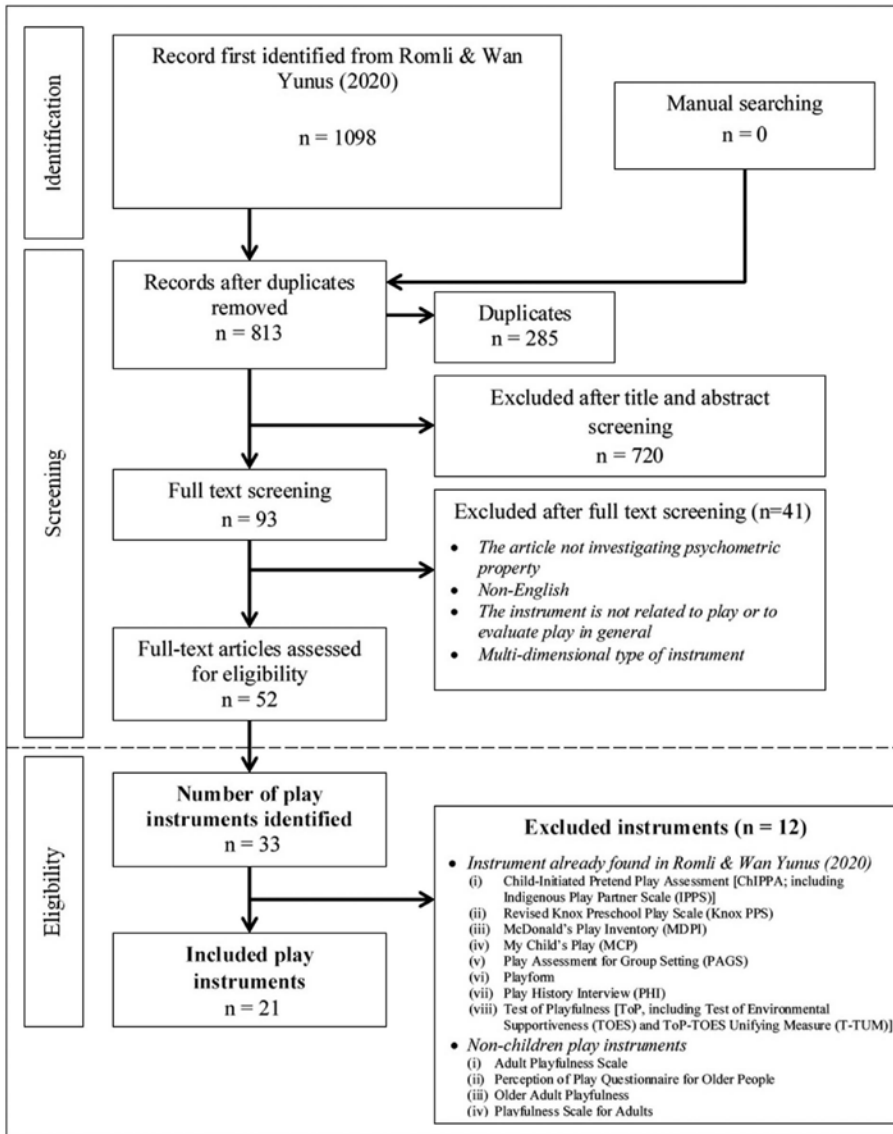
Figure 1. Prisma flowchart on screening process

Test, Singer's Observational Play Instructions and Imaginative Play Predisposition Interview, Transdisciplinary Play-Based Assessment—Child Development Resources, The Play Checklist, and Affect in Play Scale). Still other instruments are reported in journals or need to be obtained by contacting authors. We men-

tion the Penn Interactive Peer Play Scale for research purposes only, but other assessments can be used in practice.

The majority (n=18) of the instruments were originally developed for use with children. Only a limited number (n=3), can be reasonably used for adolescents, such as the Children's Active Play Imaginary Questionnaire, Fair Play Questionnaire, and The Social Play Record. Most instruments (n=19) were originally developed for typical children, and only a small number (n=2), such as CPTI and The Test of Pretend Play, were developed for children with special needs (e.g., children with developmental disabilities, children in psychotherapy such as conduct disorders, avoidant disorder, separation anxiety disorders, stress disorders, or physical child abuse). Over half (n=14) of the instruments require no or minimal training (e.g., Affect in Play Scale, Children's Active Play Imagery Questionnaire, Children's Developmental Play Instrument, Children's Playfulness Scale, Enjoyment of Lunchtime Play Survey Cards, Fair Play Questionnaire, Howes Peer Play Scale, Lunzer's Scale of Play Behaviour, Parten's Social Play Hierarchy, Penn Interactive Peer Play Scale, Play Behaviour Observation System, Play Performance Scale, Singer's Observational Play Instructions and Imaginative Play Predisposition Interview, and The Test of Pretend Play). In terms of duration, less than half (n=9) of the instruments can be completed in under thirty minutes (e.g., Affect in Play Scale, Children's Active Play Imagery Questionnaire, Children's Developmental Play Instrument, Children's Playfulness Scale, Enjoyment of Lunchtime Play Survey Cards, Fair Play Questionnaire, Penn Interactive Peer Play Scale, Play Performance Scale, Singer's Observational Play Instructions and Imaginative Play Predisposition Interview, and Symbolic Play Test).

Several citations indicate that, based on popularity, older instruments tend to rank higher than more recent ones (figure 3). The oldest instrument was developed in 1932 with 2,470 citations reported in Google Scholar, while the most recent instrument was developed in 2019, with only 18 citations. The least cited had only 2 citations on several instruments. However, having the highest numbers of citations does not reflect the highest quality of the cited instruments. Such a determination needs not only to reflect the quality but also the design robustness and the psychometric evidence of the assessment (figure 4). We gave all authors guidelines and a briefing before they rated the instruments. We based the rating description of this article given to the authors in reviews by Francis and his colleagues (Francis et al. 2016). The highest score using the quality tool of Francis and colleagues went to the Children's Active Play

Figure 2. Characteristics of the included instruments

Instrument	Purpose	Population	Scoring procedure		Duration	Assessor	Training requirement
			Description	**Scoring**			
Affect in Play Scale Version available: (i) School-age (ii) Preschool (iii) Brief Rating	To assess cognitive (imagination, organization, elaboration, comfort) and affect aspects in the child's puppet show play.	Children aged 6 – 10 years old (Pre-school and school children).	The school-age and preschool version requires videotaping but not for the brief version.	All domains except for effect are scored on a 5-point Likert scale based on observation. The effect has 12 categories and is scored by counting the verbal and non-verbal affect expressions.	5 minutes	Any professionals on observation	No formal training is available however, any training is encouraged.
Children's Active Play Imagery Questionnaire	An 11-item instrument to assess three factors (fun, social and capability) and identify the type of imagination used.	Children aged 7 – 14 years old.	NIL	Each item is scored on a 5-point Likert scale. The total score is calculated by adding the score of each item on each domain.	5 – 10 minutes	Self-rating on the questionnaire	No training is required.
Children's Developmental Play Instrument (CDPI)	To observe the developmental aspect of typical children on play contributing elements and coping adaptation. Functional aspects evaluated are level of play engagement, level of symbolic play, and play style.	Children aged 1 – 9 years old.	The instrument has a three-stage procedure with 7 domains comprised of 22 questions in total.	Most of the items are scored on a 2-point Likert scale. Has a 5-point Likert scale version.	10 minutes	Rating by observation of evaluator	No training is required.

Figure 2, continued

Instrument	Purpose	Population	Structure	Scoring	Time	Rated by	Training
Children's Play Therapy Instrument (CPTI)	To observe the developmental aspect of children with a clinical diagnosis on play-contributing elements and coping adaptation. Functional aspects evaluated are level of play engagement, level of symbolic play, and play style.	Children with clinical diagnosis.	The instrument has a three-level procedure with 12 categories comprised of 35 items in total.	Each item is scored on a 5-point Likert scale.	20 – 50 minutes	Rating by observation of play therapist	Needs to be trained clinical psychologist. Formal training on the use of CPTI is not required but encouraged.
Children's Playfulness Scale	The instrument is to measure the degree of playfulness among children.	Children aged 27 – 68 months old	NIL	A 5-point Likert scale is scored on 23 descriptive-statement items by the caregiver (i.e. parent) or teacher. Some items are inverted coding to get consistent. The total score is calculated by summing each item's score. A higher score indicates a better outcome.	15 – 20 minutes	Proxy-rating by caregiver or teacher	No training is required. However, a 16-hour of training improves the outcome.
Enjoyment of Lunchtime Play Survey Cards	To measure children's enjoyment in school play activities during school recess. The instrument focus on a socioecological model that emphasizes person-environment fit.	Children aged 8 – 12 years old.	The instrument has 39 items in 3 domains: (6 categories = 20 items), interpersonal (1 category = 2 items) and physical environment and policy (5 categories = 17 items.	Each item is scored on a 5-point Likert scale. The score is calculated by adding the individual score and converted to the mean for each category.	≈10 minutes	Self-rated by the children with the guidance of parent/teacher	No training is required.

Figure 2, continued

Fair Play Questionnaire	To assess the level of social and moral behaviour of a student in sports activity.	Children aged 10 – 15 years old.	The instrument consists of two domains: two pro-social sub-scales (respect towards teammates, respect for convention) and two anti-social sub-scales (gamesmanship, cheating) comprising 15 items.	Scored on a 5-point Likert scale. Three ways to calculate the total score: i) means on each sub-scales, ii) means on each domain, or iii) total score from all items in which several items need to be reversed the score.	5 – 7 minutes	Self-rating by the children.	No training required
Howes Peer Play Scale	The method indicates the degree of play continuum on children's social behaviour during play.	Children aged 18 – 43 months old.	Using similar to Parten's 6 types of play	Assessed against a 4-point scale of no social interaction, through contingent social interaction to reciprocal and complementary interaction.	Based on several subjective observation	Proxy-rating by observer/ evaluators	No training is required.
Lunzer's Scale of Play Behaviour	To evaluate the child's capacity for the organization of play.	Children aged 2 – 6 years or equivalent mental age.	The assessment has two domains, namely, adaptiveness and integration. It is assessed on observation of unstructured and structured play sessions	Scored using Lunzer's five-point scale (1-5, with five representing the most mature form of play behaviour) for each domain.	Subjective observation of a minimum of one hour ranged over at least four days.	Proxy-rating by observer/ evaluators	No training is required. Reading from the article.

Figure 2, continued

Mature Make-Believe Play Observational Instrument	The instrument aims to identify implementation features and causal evaluations of classroom-based play interventions.	Children aged 36 – 60 months old.	The instrument has two dimensions with respective sub-dimensions: (i) child (child-related props, child meta play, play interaction, children's role-playing, child role speech and communication), and (ii) adult (centre management, planned play time, teacher intervention). It has 23 items.	Scored against 4 levels, with the total score summed up from all items.	≈ 51.18 minutes	Proxy-rating by teachers.	Requires training and manual reading.
Parten's Social Play Hierarchy	The method is to evaluate the continuum's absence or presence on children's social behaviour during play.	Children aged < 2 to 11 years old.	A simple perceived percentage denotation on play performance showed by the child on six play types: (i) Unoccupied behaviour, (ii) Solitary play, (iii) Onlooker behaviour, (iv) Parallel play, (v) Associative play, (vi) Cooperative or Organized Supplementary play.	The number of minute samples necessary to insure an unvarying picture of social participation was determined by assigning arbitrary weights to each category, summing the scores for the odd and even days for individual children, and running correlations.	Several subjective observations using the one-minute sampling observational method.	Proxy-rating by teachers.	No training is required.

Figure 2, continued

Instrument	Description	Age	Items	Scoring	Time	Rater	Training
Penn Interactive Peer Play Scale Version available: (i) Parent (ii) Teacher	The instrument is for research purpose-only and to observe the child's competencies and needs within peer play behaviour. The parent version is to assess play in the home and neighbourhood, whereas the teacher version examines play in the classroom and at school.	Aged was not specified but has been reportedly used for children aged 3 to 13 years old.	The instrument has 32 items. There are three dimensions available: (i) play interaction, (ii) play disruption, and (iii) play disconnection.	Rated on a 4-point Likert scale.	5 – 10 minutes	Proxy-rating by parents, teachers, or other who is close to the child (i.e., caregiver)	No training is required. However, having a brief training can improve the rating quality.
Play Behaviour Observation System	The instrument is to evaluate how a child is socially responding during play.	Children aged 3 – 5 years old.	The instrument has 20 items. It can be denoted into housekeeping, blocks, manipulatives, and circle categories.	Scored against 4 hierarchy levels of participation (solitary, parallel, associative, cooperative) with a range of total scores from 22 to 108 for overall play.	10 minutes rating after around 4 times 10 minutes of observation.	Personnel trained in psychology.	No training is required. However, having a brief training can improve the rating quality.
Play Performance Scale	The instrument is a generic scale based on the perception of an individual's performance in play and the restriction present.	Children aged from birth to 16 years old.	The instrument is a single-scale rating of 0 to 100 (with a 10-point interval) against the child's actual age-appropriate play performance.	The score is categorized into three outcomes (i) Moderate to severe restriction (0 – 40), (ii) mild to moderate restriction (50 – 70), and (iii) normal (80 - 100).	≤5 minutes	Proxy-rating by parents	No training required

Figure 2, continued

Instrument	Purpose	Age	Components	Scoring	Time	Rater	Training
Revised Class Play	Measure of peer reputation designed to improve the assessment of social competence as well as the psychometric properties of class play method.	School-age children	The instruments consist of three factors: (i) sociability–leadership, (ii) aggressive–disruptive, (iii) sensitive–isolated.	The scoring is based on nominations received from class peers and sorted into the factors for score calculation.	Not specified.	Peers rating with assistance from professionals.	Require manual referring.
Singer's Observational Play Instructions and Imaginative Play Predisposition Interview	The method is to observe the quality of pretend play in a child's play activity.	Children aged 3 – 5 years old.	NIL	A 5-level score on the presence of imaginative play is denoted after a 10-minute observation (1=lack any imaginative play, 2=occasional brief reference imagination, 3=moderate pretend play (more than 3 minutes involving pretend play), 4=substantial amount of pretence, simulated vocalizations and role-playing, 5=original, multifaceted pretence).	10-minute observation of free play, both indoor and outdoor.	Proxy-rating by professionals.	No training is required. However, it requires purchase and manual reading
Symbolic Play Test	The instrument asses for receptive and expressive language development on the early concept formation and symbolization.	Children aged 1 – 3 years old	The instrument contains standardized batteries for assessment purposes. The latest version is the 2nd edition.	The scoring system is based on the number of meaningful responses and connections the child is able to make. The scale was made up of 24 items	10 – 15 minutes	Proxy rating by a psychologist and speech and language therapists only.	Require purchase of the assessment kit and manual, and training.

Figure 2, continued

The Play Checklist	The instrument provides a guide on developmental play skills in sociodramatic play.	Children aged 3 – 5 years old.	The instrument has 10 independent items assessed on a hierarchy level between 3 to 7 levels, depending on the item, respectively.	The rating is based on the highest-level skills observed in each of the 10 sections.	Few sessions around 2 or 3 observations	Proxy rating by professionals, either health or non-health.	Require manual reading
The Social Play Record	The instrument assesses social play and social interaction needs. It can be used for any child, particularly with autism spectrum disorders,	Suitable for infants to adolescents (reported use for children aged 2 – 15 years old) and at any age or stage of development.	Four domains: (i) play observation, (ii) peer relationship, (iii) peer preferences, and (iv) peer friends.	The total score is 180 for four domains	Observation. Duration not specified.	Proxy rating by relevant professionals.	Require purchase of the book.
The Test of Pretend Play (aka: Warwick Symbolic Play Test)	The instrument aims to assess symbolic play competency in children on the ability to substitute, refer to an absent object, and attribute a property. The instrument involves minimal language use.	Children aged 18 months to 6 years old. Can be used for up to 8 years for children with developmental disabilities.	The instrument contains 13 materials in four sections: (i) everyday objects, (ii) doll and non-representational objects, (iii) teddy, and (iv) self. It has two observation conditions spontaneous play and structured situation.	A score of 2 was awarded for each example of original symbolic play or 1 for an imitated symbolic behaviour.	Two 20-minute periods of play observation	Proxy rating by relevant professionals.	Training is not compulsory but encouraged.

Figure 2, continued

| Transdisciplinary Play-Based Assessment Has 1st and 2nd version | To assess play development by observing the child's play with family members, peers, and professionals. | Children aged from birth - 6 years old. | The latest version (TPBA-2) has 118 items in total across four aspects: (i) sensorimotor, (ii) emotional and social, (iii) communication, and (iv) cognitive. | Requires around 3 weeks or less for observation. | 60 – 90 minutes. | Proxy rating by professionals, either health or non-health. | 10 hours cumulative or the equivalent of training. |

Figure 3. Citation of the instruments based on Google Scholar

Instrument	Original citation	# of citation
Parten's Social Play Hierarchy	Parten, M. 1932. "Social participation among preschool children." *Journal of Abnormal and Social Psychology* 27, no. 3: 243–269. doi:10.1037/h0074524.	2470
Revised Class Play	Masten, A. S., P. Morison, and D. S. Pellegrini. 1985. "A revised class play method of peer assessment." *Developmental Psychology* 21, no. 3: 523–533. doi:10.1037/0012-1649.21.3.523	1144
Singer's Observational Play Instructions and Imaginative Play Predisposition Interview	Singer, J. L. 1973. *The child's world of make-believe: Experimental studies of imaginative play*. Academic Press.	999
Transdisciplinary Play-Based Assessment – Child Development Resources	Linder, T. W. 1993. *Transdisciplinary play-based assessment: A functional approach to working with young children*. Rev. Paul H Brookes Publishing.	557
Children's Playfulness Scale	Barnett, L. A. 1991. "The playful child: Measurement of a disposition to play." *Play and Culture* 4, no. 1: 51-74.	421
Penn Interactive Peer Play Scale Version available: (i) Parent (ii) Teacher	Fantuzzo, J., B. Sutton-Smith, K. C. Coolahan, P. H. Manz, S. Canning, and D. Debnam. 1995. "Assessment of preschool play interaction behaviors in young low-income children: Penn Interactive Peer Play Scale." *Early Childhood Research Quarterly* 10, no. 1: 105–120. doi:10.1016/0885-2006(95)90028-4	346
Howes Peer Play Scale	Howes, C. 1980. "Peer play scale as an index of complexity of peer interaction." *Developmental Psychology* 16, no. 4: 371–372. doi:10.1037/0012-1649.16.4.371	261
Play Performance Scale	Lansky, S. B., M. A. List, L. L. Lansky, C. Ritter-Sterr, and D. R. Miller. 1987. "The measurement of performance in childhood cancer patients." *Cancer* 60, no. 7: 1651–1656. doi:10.1002/1097-0142(19871001)60:7<1651::aid-cncr2820600738>3.0.co;2-j	141
Play Behaviour Observation System	Farmer-Dougan, V., and T. Kaszuba. 1999. "Reliability and Validity of Play-based Observations: relationship between the PLAY behaviour observation system and standardized measures of cognitive and social skills." *Educational Psychology* 19, no. 4: 429–440. doi:10.1080/0144341990190404	100
Lunzer's Scale of Play Behaviour	Hulme, I. and E. A. Lunzer, 1966. "Play, language and reasoning in subnormal children." *Journal of Child Psychology and Psychiatry* 7: 107-123. doi:10.1111/j.1469-7610.1966.tb02168.x	86
Symbolic Play Test	Gould, J. 1986. "The Lowe and Costello Symbolic Play Test in socially impaired children." *Journal of Autism and Developmental Disorders* 16, no. 2: 199–213. https://doi.org/10.1007/BF01531730	84
The test of Pretend Play (aka: Warwick Symbolic Play Test)	Lewis, V., J. Boucher, and A. Astell. 1992. "The assessment of symbolic play in young children: A prototype test." *International Journal of Language and Communication Disorders* 27, no. 3: 231-245. https://doi.org/10.3109/13682829209029423	51

Figure 3, continued

Children's Play Therapy Instrument	Chazan, S. E. 2000. "Using the children's play therapy instrument (CPTI) to measure the development of play in simultaneous treatment: A case study." *Infant Mental Health Journal* 21, no. 3: 211-221. doi:10.1002/1097-0355(200007)21:3<211::AID-IMHJ6>3.0.CO;2-H	40
Affect in Play Scale Version available: (i) School-age (ii) Preschool (iii) Brief Rating	Russ, S. W., L. Niec, and A. Kaugars. 2000. "Play assessment of affect: the Affect in Play Scale." In *Play diagnosis and assessment* (2nd edition). Edited by K. Gitlin-Weiner, A. Sandgrund, and C. Schaefer. 722–748. New York: Wiley. pp. 722–748.	32
The Social Play Record	White, C. 2006. *The social play record: A toolkit for assessing and developing social play from Infancy to adolescence.* Jessica Kingsley Publishers.	28
Children's Active Play Imagery Questionnaire	Cooke, L., K. Munroe-Chandler, C. Hall, D. Tobin and M. Guerrero. 2014. "Development of the children's active play imagery questionnaire." *Journal of Sports Sciences* 32, no. 9: 860–869. doi:10.1080/02640414.2013.865250	24
Enjoyment of Lunchtime Play Survey Cards	Hyndman, B. P., A. C. Benson, S. Ullah, C. F. Finch, and A. Telford. 2014. "Children's Enjoyment of Play during School Lunchtime Breaks: An Examination of Intraday and Interday Reliability." *Journal of Physical Activity and Health* 11, no. 1: 109–117. doi:10.1123/jpah.2011-0200	24
Mature Make-Believe Play Observational Instrument	Germeroth, C., E. Bodrova, C. Day-Hess, J. Barker, J. Sarama, D. H. Clements, and C. Layzer. 2019. "Play It High, Play It Low: Examining the Reliability and Validity of a New Observation Tool to Measure Children's Make-Believe Play." *American Journal of Play* 11, no. 2: 183-221. https://www.museumofplay.org/app/uploads/2022/01/11-2-Article-3.pdf	18
Children's Developmental Play Instrument	Chazan, S. E. 2009. "Observing play activity: The Children's Developmental Play Instrument (CDPI) with reliability studies." *Child Indicators Research* 2: 417–436. doi:10.1007/s12187-009-9043-9	17
Fair Play Questionnaire	Hassandra, M., M. Goudas, A. Hatzigeorgiadis, and Y. Theodorakis. 2002. "Development of a questionnaire assessing fair play in elementary school physical education." *Athlitiki Psychologia* 13: 105–126.	14
The Play Checklist	Heidemann, S., and D. Hewitt. 2014. *When Play Isn't Fun: Helping Children Resolve Play Conflicts*. Redleaf Press.	2

Figure 4. Evidence on psychometric properties of the play instruments

Instrument	Conceptual Model			Content Validity			Reliability			Construct Validity			Interpretation & Scoring			Burden & Presentation			Total
	Construct defined	Target population	Expected subscales	Patient devised items	Content experts	Description of item	Reliability tested	Coefficient adequate	Justification of	Longitudinal validity	Convergent validity	Known-group	Plan of scoring	Plan for missing data	Scaling described	Length reasonable	Literacy level	Items viewable	
Affect in Play Scale – Preschool	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1	14
Children's Active Play Imagery Questionnaire	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	16
Children's Developmental Play Instrument	1	1	1	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	15
Children's Play Therapy Instrument	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	15
Children's Playfulness Scale	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	16
Enjoyment of Lunchtime Play Survey Cards	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	1	0	0	8
Fair Play Questionnaire	1	1	1	1	1	0	1	1	1	0	0	1	1	0	1	1	0	0	11
Howes Peer Play Scale	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	6
Lunzer's Scale of Play Behaviour	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	7
Mature Make-Believe Play Observational Instrument	1	0	1	0	1	1	1	1	1	0	0	1	1	0	1	0	1	1	12
Parten's Social Play Hierarchy	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	5
Penn Interactive Peer Play Scale	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	9

Figure 4, continued

Instrument	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
Play Behaviour Observation System	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	**6**
Play Performance Scale	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	**4**
Revised Class Play	1	1	1	0	1	1	1	0	1	1	0	1	1	0	1	1	0	0	0	**12**
Singer's Observational Play Instructions and Imaginative Play Predisposition Interview	1	0	0	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	**9**
Symbolic Play Test	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	**2**
The Play Checklist	1	1	1	0	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	**11**
The Social Play Record	1	1	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	0	0	**11**
The test of Pretend Play (aka: Warwick Symbolic Play Test)	1	1	0	1	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	**12**
Transdisciplinary Play-Based Assessment	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	**13**

Imaginary questionnaire, developed in 2014 (Score=16), and the Children's Playfulness Scale, developed in 1991 (Score=16). The lowest quality score went to the Symbolic Play Test, developed in 1986 (Score=2), and the second-lowest score to the Play Performance Scale, developed in 1985 (Score=4). Not all instruments have undergone validity and reliability testing for psychometric properties. A detailed description of the psychometric properties of the selected assessments appear in figure 5, focusing on validity and reliability.

Discussion

We present a total of twenty-one assessments in this article. The assessments we found vary, indicating many of them are suitable for typical children and some are for children with special needs. A considerable number of instruments are generic in nature, take only a short time to complete, and can be administered by individuals without specialized training or expertise. These instruments still offer satisfactory psychometric evidence. In this regard, it is beneficial to promote the use of standardized play assessment in practice (Cubas and Levratto 2019; Kuhaneck et al. 2013; Romli and Wan Yunus 2020). Nevertheless, due to diverse applicability and usability of such instruments, the purpose of this review is not prescriptive but to make readers aware of both the scope and purpose of the instruments we discovered. Each measurement has its own uniqueness and purpose, which is focused on a particular aspect of attention, such as on cognitive-only or physical-only observation. Our review provides the comprehensive information readers need to decide whether a measure is auspiciously tailored for their use.

Because this article and Romli and Wan Yunus (2020) are the same project, we considered the additional eight instruments from Romli and Wan Yunus, comparing both with previous reviews. Sixteen instruments (fourteen assessments from this article and two from Romli and Wan Yunus) are new and not identified previously (Brooke 2004; Burgarelli et al. 2018; Cubas and Levratto 2019; O'Grady and Dusing 2015). We also discuss the 80 percent (n=8/10) of instruments found by Cubas and Levratto. The two instruments not covered by them are the Cicchetti Scale and the Casby Scale, which are not available in English. The comparison is reasonable because Cubas and Levratto also seek an objective search for play-specific assessments. Only 14 percent (n=1/7) of assessments identified in the systematic review by O'Grady and Dusing were covered

Figure 5. Description of the validity and reliability of the included assessments

Instrument	Validity	Reliability
Affect in Play Scale	The preschool version is highly correlated with emotional understanding and with teachers-rating on creativity on a laboratory task, teacher-rated affect expression, imagination, pleasure during play, teacher-rated social competence, and general adaptation, and teacher-rated intensity of positive emotions in daily activities, prosocial classroom activities, low physical aggression in the classroom. The brief version correlated positively to hopeful thinking and problem-solving and related negatively to teacher-reported anxiety.	Inter-rater: ICC=>0.90 (preschool version) ICC=0.82-0.96 (brief version), Low (school-age version). Internal consistency: Split-half=0.82 (preschool version.)
Children's Active Play Imagery Questionnaire	Factorial validity Comparative Fit Index = 0.95; Normative Fit Index = 60.92; Tucker-Lewis Index = 0.93, and RootMean-Square Error of Approximation = 0.07.	Internal consistency: α=0.73–0.82
Children's Developmental Play Instrument	Construct validity with factor analysis resulted in three play style clusters: adaptive, impulsive/disorganized, and conflicted/inhibited.	Inter-rater reliabilities: k = 0.71–0.80
Children's Play Therapy Instrument	Good convergent and predictive validity in relation to internalizing and externalizing problems and discriminant validity in differentiating traumatic versus typical play characteristics	Good interrater reliability. The measure is sensitive to changes in psychotherapy.
Children's Playfulness Scale	Factor analyses comprised five dimensions: "physical spontaneity"; "social spontaneity", "cognitive spontaneity"; "manifest joy"; "sense of humour", with 87% accounted for common variance with > 0.55-factor loading.	Inter-rater: 0.922, Test-retests: 0.958 – 0.971; Internal consistency: α = 0.88; Ω = 0.87 – 0.95
Enjoyment of Lunchtime Play Survey Cards	Children reported little concern or difficulty using the small survey cards. The survey cards' suitability for children under 10 years was acceptable based on feedback from elementary school teachers after the initial pilot study. The face validity of the small survey card was reviewed by 5 physical activity experts with experience in the development of self-report measures. Readability was also assessed by conducting a pilot test with 15 children aged 8–12 years.	Intraday kappa values ranged from: Fair (0.31) to substantial (0.75) within each of the 5 days (k = 0.41). In comparison "expected" (0.09 – 0.40; median 0.30) and "actual" (0.05–0.46; median 0.28) interday enjoyment of lunchtime play displayed low reliability.
Fair Play Questionnaire	Factor analysis on the Greek population aged 10 to 12 years has demonstrated the four-dimensional structure of the instrument. Internal consistency of the subscales has been supported (α = 0.62 – 0.81). This scale was completed three times, in pre-, post- and follow-up measures by both groups.	Internal consistency coefficient; Between 0.66 and 0.89.

Figure 5, continued

Howes Peer Play Scale	Measures complexity of social play. Four original was used; Parallel play k = 0.93; Parallel aware play k = 0.92; Simple social play k = 0.94; Complementary and reciprocal play k = 0.94.	Cronbach's alpha ranged from 0.66 to 0.89. Test-retest reliability: Reported at 0.91 over 4 weeks and 0.34 over 1 year
Lunzer's scale of play behaviour	Nil	Inter-rater reliabilities: Adaptiveness in Play 0.91 – 0.96, Integration in Play 0.90 – 0.91.
Mature Make-Believe Play Observational Instrument	Tested predictive validity on how to play related to children's; Self-regulatory, most self-regulation measures showed high association with MPOT; Math, 32% of the variance in TEAM scores was explained by classroom level variance in a no-intercept model (ICC = 0.32); Literacy performance, mature play predicts some emerging literacy skills.	Inter-rater reliability: In classroom-based field testing (α = 0.909) is high. The dimensions and cumulative scores were found to have moderate correlations with self-regulation outcome measures. ICC ranged from 0.11 (pencil-tap) to 0.18 (HTKS), with 11% to 18% of the variance in child scores. ICCs for each variable ranged from 0.05 (RBS-NA) to 0.45 (PALS letter sounds)
Parten's Social Play Hierarchy	Nil	Correlation coefficient: Ten even-day observations correlated with 10 odd-day observations of social participation $r = 0.79$. With 20 even and odd day samples the correlation coefficient obtained was $r = 0.90$. The correlations of 10 even and 10 odd-day scores of leadership observations $r = 0.39$. Twenty scores of alternate days correlated $r = 0.44$; while thirty sample of even and odd days yielded a correlation coefficient of $r = 0.73$.
Penn Interactive Peer Play Scale	Convergent validity was confirmed. PIPPS Play an Interaction factor, and the SSRS's Interpersonal Skills, Verbal Assertion, and Self Control factors are significantly correlated ($p<0.001$). PIPPS Disconnection factor moderately correlates with SSRS Externalizing and Internalizing Behavior Problems ($p<0.01$). PIIPS Disruption factor and SSRS Externalizing Behavior Problems correlated moderately ($p<0.001$). Divergent validity was also confirmed.	Interrater reliability: High correlation of 0.88, $p<0.001$. Internal consistency: α = 0.90 (high) for the Play Disruption factor, α = 0.89 (high) for the Play Disconnection factor α = 0.90 (high) for the Play Interaction factor.

Figure 5, continued

Instrument	Validity	Reliability
Play Behaviour Observation System	Concurrent validity was established with only one single model generated. The total play score significantly correlated with the overall Battelle Developmental Inventory (BDI) score ($p = 0.002$). The play score obtained during circle time was positively correlated with TSS ($p = 0.007$) and TPB ($p = 0.012$) of the SSRS-T.	High internal reliability with the interobserver agreement (IOA) varied between 80% and 100% for all observations, with a mean of 92% (SD = 8.4).
Play Performance Scale	Correlation coefficients in the group of 18 inpatients and nurses' ratings were highly related to parent play-performance ratings ($r = 0.75$, $P < 0.001$). Interviewer ratings were also highly correlated with parent ratings ($r = 0.92$, $p < 0.0001$).	Interrater reliability: Good between parents, $r = 0.71$. Further calculations revealed that parents were in complete agreement two-thirds (67%) of the time.
Revised Class Play	Good validity. Construct validity based on factor analysis in three dimensions: All positive items loaded substantially on the first factor labelled as "Sociability-Leadership". The negative items were divided into two factors labelled as "Aggressive-Disruptive' and "Sensitive-Isolated".	Good reliability. Internal consistency: Positive Scale $\alpha = 0.93 - 0.95$; Disruptive Scale $\alpha = 0.90 - 0.93$; Isolation Scale $\alpha = 0.81 - 0.85$.
Singer's Observational Play Instructions and Imaginative Play Predisposition Interview	Nil	Inter-rater reliability: Imaginativeness $k= 0.94 - 0.96$, Emotional Affect $k= 0.92 - 0.96$, Concentration $k= 0.92 - 0.92$ Aggression $k= 0.96 - 1.00$.
Symbolic Play Test	Concurrent validity reported significant correlations between; verbal and non-verbal IQ and the SPT: 0.506 and 0.461, respectively; verbal and nonverbal IQ and the 'imaginativeness of play' rating: 0.52 and 0.356, respectively.	The inter-rater: Correlation for this measure is 0.92; Inter-observer correlation is 0.82.
The Play Checklist	Nil	There is no evidence that the measure's reliability has been tested, although detailed administrative guidelines are provided in the main text.
The Social Play Record	There is no evidence of statistical tests of validity.	No high-quality evidence that the measure's reliability has been tested in the Social Play Record Toolkit publication or any other identified work.
The test of Pretend Play (aka: Warwick Symbolic Play Test)	Concurrent validity was tested by correlating the test scores, language ability, and non-ability. Construct validity was reported at 0.86.	Test-retest reliability: Reported as 0.87 (in a sample of 40 children). Internal consistency: Ranged from 0.55 to 0.94. Inter-rater reliability: Considered as good agreement with scores of 0.68.
Transdisciplinary Play-Based Assessment – Child Development Resources	Strong concurrent correlation with Peabody Developmental Motor Scales-2 for developmental age at baseline for gross- ($r = 0.947$) and fine-motor ($r = 0.918$) domains also also at 6 months later for the gross- ($r = 0.937$) and fine-motor ($r = 0.967$) domains. Responsiveness was demonstrated in early intervention ($p<0.01$). The correlation coefficient between staff and parents for standardized assessments was 0.67 ($p = 0.001$) and 0.70 ($p = 0.001$) for play-based assessments.	Test-retest reliability is well supported for all developmental domains—high interrater reliability for language, motor and combined domains sufficient for educational placement purposes.

p = p-value; r = correlation coefficient; ICC = intraclass correlation; α; alpha; Ω = omega; k = kappa

in our reviews. This is expected because four of the assessments in O'Grady and Dusing are nonplay assessments designed to measure motor and cognitive skills, although some of the items in the instruments have elements of play. Only one of the three assessments specified by Brooke (2004) (i.e., Developmental Play Assessment) focus solely on assessing play. This could explain why none of the assessments found by Brooke was present in our review. Around 52 percent (n=15 of 29) of the assessments listed by Bulgarelli and his colleagues (Bulgarelli et al. 2018) were also identified in our review. Burgarelli and his colleagues included several assessments that set play as part of its construct. For example, the Children's Assessment of Participation and Enjoyment and Preferences for Activities of Children assesses general participation in functional performance, and play is asked about in just some of its items. Hence, some assessments were not identified in our search, which is reasonable because we focus on assessments that consider only play. Assessments found by our project manage to focus only on play assessments, not play-based or nonplay assessments.

Frequently cited assessment does not necessarily yield better quality or practice. For example, Parten's Social Play Hierarchy, despite being the most cited, has been criticized for having limited psychometric evidence and for being cited as a basis for knowledge. Prior to the 1990s, the majority of assessments were poorly designed, subjective, and lacked analytic outcomes. In contrast, more recently developed assessments are statistically constructed and have undertaken extensive research. Consequently, researchers and practitioners should consider the feasibility and evidence available when selecting an assessment.

Observation is one of the more frequent methods used by practitioners (Bulgarelli and Stancheva-Popkostadinova 2018; Md Yakup and Dahlan 2022). Although subjective observation is a powerful method for providing rich information, objective measurement is necessary for evidence-based purposes (Skinner and Turner-Stokes 2006; Stapleton and McBrearty 2009). Our review identified a number of assessments available to elevate the observational practice with a minimal administrative burden, including the Howes Peer Play Scale, Children's Developmental Play Instrument, and Transdisciplinary Play-Based Assessment—Child Development Resources.

Limitations

We do not identify several frequently mentioned assessments. For instance, the

Play in Early Childhood Evaluation System (PIECES) (Kelly-Vance and Ryalls 2005) and Play Assessment Scale are the most frequently used and with good evidence available. But they are not included because they do not meet our defined criteria—which they have published in gray literature. Some assessment tools are not captured in our search. As a result we missed, for example, the System for Observing Children's Activity and Relationships During Play (SOCARP) (Ridgers et al. 2010).

Moreover, because the review search we conducted is relatively dated, several recently published assessments, such as the Children Play Scale (Dodd, Nesbit, and Maratchi 2021), may not have been yet retrievable in our search. We were unable to conduct an updated search due to limited manpower and limited funding. In addition, although our search was systematic, it is possible we missed available assessments because of limitations in the searching algorithm and our knowledge.

Conclusion

Practitioners should use instruments to measure what they purport to measure. As noted by Bulgarelli and Stancheva-Popkostadinova (2018), practitioners employ a number of instruments that use play to assess nonplay aspects such as cognitive and interaction (e.g., Early Start Denver Model Curriculum) or were not intended for use in play (e.g., Autism Diagnostic Tool—Revised). And some of these instruments were not evaluation tools. Md Yakup and Dahlan (2022) observed that the vast majority of practitioners do not use standardized play instruments and that they use play as a mere observation or just as a reward. One study found that practitioners used an assessment primarily because of its availability, ease of administration, and exposure during undergraduate training rather than for its assessment evidence or for searching and exploring new assessments. The use of inappropriate instruments may mislead practitioners' judgment on intervention planning. Nevertheless, this review, combined with Romli and Wan Yunus (2020), can provide a preliminary guide for practitioners interested in play to refer, select, and use standardized assessment in their work.

REFERENCES

Barnett, Lynn A. 1991. "The Playful Child: Measurement of a Disposition to Play." *Play and Culture* 4:51–74.

Brooke, Stephanie L. 2004. "Critical Review of Play Therapy Assessments." *International Journal of Play Therapy* 13:119–142. https://doi.org/10.1037/h0088893.

Bulgarelli, Daniela, Nicole Bianquin, Francesca Caprino, Paola Molina, and Sylvie Ray-Kaeser. 2018. "Review of Tools for Play and Play-Based Assessment." In *Evaluation of Children's Play: Tools and Methods*, edited by Serenella Besio, Daniela Bulgarelli, and Vaska Stancheva-Popkostadinova, 58–113. https://doi.org/10.1515/9783110610604-005.

Bulgarelli, Daniela and Vaska Stancheva-Popkostadinova. 2018. "Play Assessment Tools and Methodologies: The View of Practitioners." In *Evaluation of Children's Play: Tools and Methods*, edited by Serenella Besio, Daniela Bulgarelli, and Vaska Stancheva-Popkostadinova, 114–35. https://doi.org/10.1515/9783110610604-006.

Bundy, Anita. 2010. "Evidence to Practice Commentary: Beware the Traps of Play Assessment." *Physical & Occupational Therapy in Pediatrics* 30:98–100. https://doi.org/10.3109/01942631003622723.

Chazan, Saralea E. 2000. "Using the Children's Play Therapy Instrument (CPTI) to Measure the Development of Play in Simultaneous Treatment: A Case Study." *Infant Mental Health Journal* 21:211–21. https://doi.org/10.1002/1097-0355(200007)21:3<211::AID-IMHJ6>3.0.CO;2-H.

———. 2009. "Observing Play Activity: The Children's Developmental Play Instrument (CDPI) with Reliability Studies." *Child Indicators Research* 2:417–36. https://doi.org/10.1007/s12187-009-9043-9.

Cooke, Lisa, Krista Munroe-Chandler, Craig Hall, Danielle Tobin, and Michelle Guerrero. 2014. "Development of the Children's Active Play Imagery Questionnaire." *Journal of Sports Sciences* 32:860–69. https://doi.org/10.1080/02640414.2013.865250.

Cubas, Marta Barcenilla, and Valeria Levratto. 2019. "Evaluación Psicopedagógica Basada en el Juego en Educación Infantil: Un Análisis Comparativo Entre Instrumentos." [Psychopedagogical Evaluation Based on Play in Early Childhood Education: A Comparative Analysis among Instruments]." *Educação e Pesquisa* 45: e203634. https://doi.org/10.1590/s1678-4634201945203634.

Dankiw, Kylie A., Margarita D. Tsiros, Katherine L. Baldock, and Saravana Kumar. 2020. "The Impacts of Unstructured Nature Play on Health in Early Childhood Development: A Systematic Review." *PLoS One* 15: e0229006. https://doi.org/10.1371/journal.pone.0229006.

Dijkstra-de Neijs, Leanne, Chanel Tisseur, Laura. A. Kluwen, Ina A. van Berckelaer-Onnes, Hanna Swaab, and Wietske A. Ester. 2021. "Effectivity of Play-Based Interventions in Children with Autism Spectrum Disorder and Their Parents: A

Systematic Review." *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-021-05357-2.

Dodd, Helen F., Rachel J. Nesbit, and Laura R. Maratchi. 2021. "Development and Evaluation of a New Measure of Children's Play: The Children's Play Scale (CPS)." *BMC Public Health* 21. https://doi.org/10.1186/s12889-021-10812-x.

Duncan, Edward A. S., and Jennifer Murray. 2012. "The Barriers and Facilitators to Routine Outcome Measurement by Allied Health Professionals in Practice: A Systematic Review." *BMC Health Services Research* 12:1–9. https://doi.org/10.1186/1472-6963-12-96.

Fantuzzo, John, Brian Sutton-Smith, Kathleen Coyle Coolahan, Patricia Holiday Manz, Sally Canning, and Darlena Debnam. 1995. "Assessment of Preschool Play Interaction Behaviors in Young Low-Income Children: Penn Interactive Peer Play Scale." *Early Childhood Research Quarterly* 10:105–20. https://doi.org/10.1016/0885-2006(95)90028-4.

Farmer-Dougan, Valeri, and Tami Kaszuba. 1999. "Reliability and Validity of Play-Based Observations: Relationship between the PLAY Behaviour Observation System and Standardized Measures of Cognitive and Social Skills." *Educational Psychology* 19:429–40. https://doi.org/10.1080/0144341990190404.

Foulds, Kim. 2023. "Playful Perceptions: The Role of and Barriers to Play for Parents of Young Children in Diverse Global Contexts." *International Journal of Play* 12:206–23. https://doi.org/10.1080/21594937.2022.2156040.

Francis, David O., Melissa L. McPheeters, Meaghan Noud, David F. Penson, and Irene D. Feurer. 2016. "Checklist to Operationalize Measurement Characteristics of Patient-Reported Outcome Measures." *Systematic Reviews* 5:1–11. https://doi.org/10.1186/s13643-016-0307-4.

Francis, Gill, Emre Deniz, Carole Torgerson, and Umar Toseeb. 2022. "Play-Based Interventions for Mental Health: A Systematic Review and Meta-Analysis Focused on Children and Adolescents with Autism Spectrum Disorder and Developmental Language Disorder." *Autism & Developmental Language Impairments*. https://doi.org/10.1177/23969415211073118.

Germeroth, Carrie, Elena Bodrova, Crystal Day-Hess, Jane Barker, Julie Sarama, Douglas H. Clements, and Carolyn Layzer. 2019. "Play It High, Play It Low: Examining the Reliability and Validity of a New Observation Tool to Measure Children's Make-Believe Play." *American Journal of Play* 11:183–221. https://www.museumofplay.org/app/uploads/2022/01/11-2-Article-3.pdf.

Gomes, Nayra Rejane Rolim, Ednei Costa Maia, and Istvan Van Deursen Varga. 2018. "Os benefícios do brincar para a saúde das crianças: uma revisão sistemática, [The Benefits of Play for Children's Health: A Systematic Review]." *Arquivos de Ciências da Saúde* 25:47–51. https://doi.org/10.17696/2318-3691.25.2.2018.867.

Gould, Judith. 1986. "The Lowe and Costello Symbolic Play Test in Socially Impaired Children." *Journal of Autism and Developmental Disorders* 16:199–213. https://doi.org/10.1007/BF01531730.

Hassandra, Mary, Marios Goudas, Antonis Hatzigeorgiadis, and Yiannis Theodorakis. 2002. "Development of a Questionnaire Assessing Fair Play in Elementary School Physical Education." *Athlitiki Psychologia* 13:105–26.

Heidemann, Sandra, and Deborah Hewitt. 2014. *When Play Isn't Fun: Helping Children Resolve Play Conflicts*.

Howes, Carollee. 1980. "Peer Play Scale as an Index of Complexity of Peer Interaction." *Developmental Psychology* 16:371–72. https://doi.org/10.1037/0012-1649.16.4.371.

Hulme, I. and E. A. Lunzer. 1966. "Play, Language, and Reasoning in Subnormal Children." *Journal of Child Psychology and Psychiatry* 7:107–23. https://doi.org/10.1111/j.1469-7610.1966.tb02168.x.

Hyndman, Brendon P., Amanda C. Benson, Shahid Ullah, Caroline Finch, and Amanda Telford. 2014. "Children's Enjoyment of Play during School Lunchtime Breaks: An Examination of Intraday and Interday Reliability." *Journal of Physical Activity and Health* 11:109–17. https://doi.org/10.1123/jpah.2011-0200.

Kelly-Vance, Lisa, and Brigette Oliver Ryalls. 2005. "A Systematic, Reliable Approach to Play Assessment in Preschoolers." *Psychology Faculty Publications* 127. https://digitalcommons.unomaha.edu/psychfacpub/127.

Kent, Cally, Reinie Cordier, Annette Joosten, Sarah Wilkes-Gillan, Anita Bundy, and Renee Speyer. 2020. "A Systematic Review and Meta-Analysis of Interventions to Improve Play Skills in Children with Autism Spectrum Disorder." *Review Journal of Autism and Developmental Disorders* 7:91–118. https://doi.org/10.1007/s40489-019-00181-y.

Kuhaneck, Heather Miller, Kari J. Tanta, Allison Kristine Coombs, and Heather Pannone. 2013. "A Survey of Pediatric Occupational Therapists' Use of Play." *Journal of Occupational Therapy, Schools, and Early Intervention* 6:213-27. https://doi.org/10.1080/19411243.2013.850940.

Lansky, S., M. List, L. Lansky, C. Ritter-Sterr, and Denis R. Miller. 1987. "The Measurement of Performance in Childhood Cancer Patients." *Cancer* 60:1651–56. https://doi.org/10.1002/1097-0142(19871001)60:7<1651::aid-cncr2820600738>3.0.co;2-j

Lewis, Vicky, Jill Boucher, and Arlene Astell. 1992. "The Assessment of Symbolic Play in Young Children: A Prototype Test." *International Journal of Language and Communication Disorders* 27:231–45. https://doi.org/10.3109/13682829209029423.

Linder, Toni W. 1993. *Transdisciplinary Play-Based Assessment: A Functional Approach to Working with Young Children*.

Masten, Ann S., Patricia Morison, and David S. Pellegrini. 1985. "A Revised Class Play Method of Peer Assessment." *Developmental Psychology* 21:523–33. https://doi.org/10.1037/0012-1649.21.3.523.

Md Yakup, Khairul Anam, and Akehsan Dahlan. 2022. "The Malaysian Occupational Therapist Perspective on the Use of Play in Children with Disabilities." *Malaysian Journal of Medicine and Health Sciences* 18: 322–32. https://doi.org/10.47836/mjmhs18.8.41.

Novak, Iona, and Ingrid Honan. 2019. "Effectiveness of Paediatric Occupational Ther-

apy for Children with Disabilities: A Systematic Review." *Australian Occupational Therapy Journal* 66:258–73. https://doi.org/10.1111/1440-1630.12573.

O'Grady, Michael G., and Stacey C. Dusing. 2014. "Reliability and Validity of Play-Based Assessments of Motor and Cognitive Skills for Infants and Young Children: A Systematic Review." *Physical Therapy* 95:25–38. https://doi.org/10.2522/ptj.20140111.

Parten, Mildred. 1932. "Social Participation among Preschool Children." *Journal of Abnormal and Social Psychology* 27:243–69. https://doi.org/10.1037/h0074524.

Ridgers, Nicola D., Gareth Stratton, and Thomas L. McKenzie. 2010. "Reliability and Validity of the System for Observing Children's Activity and Relationships during Play (SOCARP)." *Journal of Physical Activity and Health* 7:17–25. https://doi.org/10.1123/jpah.7.1.17.

Romli, Muhammad Hibatullah, and Farahiyah Wan Yunus. 2020. "A Systematic Review on Clinimetric Properties of Play Instruments for Occupational Therapy Practice." *Occupational Therapy International*. https://doi.org/10.1155/2020/2490519.

Rosenkoetter, Ulrike, and Robyn L. Tate. 2017. "Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools." *Brain Impairment* 19:103–18. https://doi.org/10.1017/brimp.2017.29.

Russ, Sandra W., Larissa N. Niec, and Astrida S. Kaugars. 2000. "Play Assessment of Affect: The Affect in Play Scale." In *Play Diagnosis and Assessment*, 2nd ed., edited by Karen Gitlin-Weiner, Alice Sandgrund, and Charles E. Schaefer. 722–48.

Salcuni, Silvia, Claudia Mazzeschi, and Claudia Capella. 2017. "Editorial: The Role of Play in Child Assessment and Intervention." *Frontiers in Psychology* 8:1098. https://doi.org/10.3389/fpsyg.2017.01098.

Shamseer, Larissa, David Moher, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, and Lesley A. Stewart. 2015. "The PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P): Elaboration and Explanation." *BMJ*. https://doi: 10.1136/bmj.g7647.

Singer, Jerome. L. 1973. *The Child's World of Make-Believe: Experimental Studies of Imaginative Play*.

Skene, Kayleigh, Christine M. O'Farrelly, Elizabeth M. Byrne, Natalie Kirby, Eloise C. Stevens, and Paul G. Ramchandani. 2022. "Can Guidance during Play Enhance Children's Learning and Development in Educational Contexts? A Systematic Review and Meta-Analysis." *Child Development* 93:1162–80. https://doi.org/10.1111/cdev.13730.

Skinner, Aneetha, and Lynne Turner-Stokes. 2006. "The Use of Standardized Outcome Measures in Rehabilitation Centres in the UK." *Clinical Rehabilitation* 20:609–15. https://doi.org/10.1191/0269215506cr981oa.

Stagnitti, Karen. 2004. "Understanding Play: The Implications for Play Assessment." *Australian Occupational Therapy Journal* 51:3–12. https://doi.org/10.1046/j.1440-1630.2003.00387.x.

Stapleton, Tadhg, and Clara McBrearty. 2009. "Use of Standardized Assessments and

Outcome Measures among a Sample of Irish Occupational Therapists Working with Adults with Physical Disabilities." *British Journal of Occupational Therapy* 72:55–64. https://doi.org/10.1177/030802260907200203.

Unsworth, Carolyn A. 2011. "Evidence-Based Practice Depends on the Routine Use of Outcome Measures." *British Journal of Occupational Therapy* 74:209. https://doi.org/10.4276/030802211X13046730116371.

White, Chris. 2006. *The Social Play Record: A Toolkit for Assessing and Developing Social Play from Infancy to Adolescence.*