Assessing Preschool Children's Social Pretend Play Competence: An Empirical Comparison of Three Different Assessment Methods

Ann-Kathrin Jaggy[a, b], Sonja Perren[a, b] and Fabio Sticca[c]

[a]Thurgau University of Teacher Education, Kreuzlingen (Switzerland)

[b]University of Konstanz (Germany)

[c]Marie Meierhofer Children's Institute, Zurich (Switzerland)

Accepted for publication in Early Education and Development, 2020
https://dx.doi.org/10.1080/10409289.2020.1712633

Author Note

Correspondence should be addressed to Ann-Kathrin Jaggy, Thurgau University of Teacher Education and University of Konstanz, Chair of Development and Education in Early Childhood, Bärenstrasse 38, 8280 Kreuzlingen (Switzerland).

Phone: +41 (0)71 678 57 48, e-mail: ann-kathrin.jaggy@uni-konstanz.de

Assessing Preschool Children's Social Pretend Play Competence: An Empirical Comparison of Three Different Assessment Methods

Pretend play may be beneficial for young children's social development. However, empirical results to date are inconsistent and limited, which is partly due to a lack of psychometrically sound measures for children's social pretend play competence. The current study aimed to compare and validate different assessment methods for children's social pretend play competence. In total, 64 3- to 4-year-old children participated in the study (age: $M = 46.4$, $SD = 3.8$). Assessments were conducted twice, three months apart. Social pretend play competence was assessed using a standardized role play test (Tools of the Play Scale), a social pretend play situation with a peer (Dyadic Pretend Play Assessment), and a teacher report. Children's Theory of Mind, emotion understanding, and language comprehension were assessed. Educators reported on children's social-emotional skills. All three instruments showed a good factorial validity, measurement invariance and sensitivity to intra-individual change. A second-order factor of all three methods was identified. The Tools of the Play Test and the teacher report yielded good criterion validity. The second-order factor showed even better criterion validity: children with higher social pretend play competence showed higher social-cognitive skills as well as social-emotional skills. Limitations and applications of the instruments are discussed.

Keywords: social pretend play, social skills, theory of mind, emotion understanding, early childhood education, assessment methods

Assessing Preschool Children's Social Pretend Play Competence: An Empirical Comparison of Three Different Assessment Methods

There are strong practical and theoretical claims that pretend play is beneficial for young children's social development (Bergen, 2013). It has been suggested that social pretend play may have a positive impact on children's social development because it is a training ground for positive and joyful peer interactions. For instance, children are required to share, to assert themselves, to take turns, as well as to show and to interpret emotions (Smith, 2010). Pretend play also allows children to create different make-believe scenarios in which children can use pretence and symbolism to take on different roles and perspectives. It can therefore also challenge their social-cognitive abilities such as theory of mind (ToM), emotion understanding or their language development. From the theoretical perspective, we therefore might assume that there is a causal link between social pretend play and children's social-cognitive and social-emotional skills.

However, from an empirical perspective, the issue of its potential causality needs further clarification. Lillard et al. (2013) conducted a critical literature review of the association between pretend play and selected cognitive, social-cognitive, and social developmental outcomes. Regarding non-social cognitive skills, the authors concluded that associations with creativity, intelligence and conservation seem to be epiphenomena, while associations with (cognitive) problem solving appear to be non-existent. However, regarding social-cognitive (e.g. ToM) and social skills, associations with *social* pretend play might be equifinal, even though sound empirical support is still missing. This is also due to a lack of measurement tools with good psychometric properties. Sound empirical methods to assess social pretend play competence are needed for further research as well as for assessing children's development. The present

study compares and combines different social pretend play assessment methods and tests them for their factorial and criterion validity.

The elusive construct of pretend play

Since pretend play has many facets, it is hard to define and to operationalise. Most often, pretend play is defined through playing "as if" (Fein, 1981; Smith, 2010). This gives a wide scope for the operationalisation of the construct. Consequently, there have been many different ways to assess children's pretend play, ranging from assessing whether or not children have imaginary companions (Taylor, Carlson, Maring, Gerow, & Charley, 2004) to tests of children's competence in standardized play settings (Overton & Jackson, 1973). This might be one reason for inconsistent empirical findings concerning the associations between children's pretend play and their social and social-cognitive development. To investigate the role of pretend play in children's development, a more precise examination of the suggested theoretical reason why it may be beneficial for children's social and social-cognitive development and a closer inspection of the construct "pretend play" are needed.

According to Vygotsky (1978) "play creates a zone of proximal development of the child" (p. 102), whereby especially pretend play in the interaction with others – *social* pretend play – seems to be important in fostering children's social and social-cognitive skills because it helps children to understand the social world and to engage in positive peer interactions. Social pretend play actions, like other-directed pretend play, could therefore be important in children's social development. Complementary to this social perspective, Piaget (1962) focused on a more cognitive view and described the strong associations between pretend play and children's cognitive development, for example for cognitive abilities such as the awareness of meta representations. Based on Piaget's

work it may be assumed, that social-cognitive play actions like substitutions are beneficial for children's social development (Lillard, Pinkham, & Smith, 2011). However, as these play actions develop and show different degrees of maturity, Bodrova, Germeroth and Leong (2013) point out that the central factor for a positive developmental outcome is not whether children pretend, but what their level of pretend play is.

To investigate these theoretical suggestions, we need psychometrically sound measures to assess children's developmental level of social pretend play, i.e. their social pretend play competence. Based on the literature (Bodrova et al., 2013; Bodrova & Leong, 2007; Smilansky & Shefatya, 1990), we identified developmentally advanced features of pretend play in the following five facets, which integrate social and cognitive components of play: *Decentration* describes a shift from self- to other-directed play actions. *Decontextualization* develops from using realistic object imitations to using the substitution of objects. *Role-taking* changes from simple dressing-up actions to role-conforming play behavior over a longer period of play. *Planning* shifts from single planned actions to meta-planned actions. Finally, *sequencing* develops from non-related single play actions to the flexible application of play scripts. However, most of the common instruments focus on single aspects of pretend play, such as the ability to substitute, and not on a more comprehensive construct of pretend play. This article therefore focuses on methods for assessing social pretend play competence as a multifaceted concept, which we define as play with an interaction partner containing the developmentally advanced features of pretend play encompassing the five mentioned social and social-cognitive facets.

Measures for assessing pretend play

There are only a few methodologically sound instruments to assess pretend play in early childhood. They vary in operationalisation, structure, and procedure. One such instrument is the Child-Initiated Pretend Play Assessment (ChiPPA; Stagnitti & Unsworth, 2004; Stagnitti, Unsworth, & Rodger, 2000), a standardized and systematic observation instrument that assesses children's pretend play competence in a simulated play setting over 30 minutes with an examiner (18 minutes play for children up to three years of age). The examiner creates a play space, using a nondirective approach to encourage the child to play by presenting conventional toys with clear functions, like miniature animals, in the first half, and unstructured materials like blankets or sticks without a clear function in the second half. During the play, the examiner rates the child's quality of pretend play actions, ability to substitute objects, and imitation of modelled actions. Using the ChiPPA, Uren and Stagnitti (2009) found positive associations between the quality and object imitation scores of 4- to 7-year-olds and their classroom involvement as well as peer play interaction scores. Furthermore, they found negative associations between the quality of the pretend play and social disruption scores. Nevertheless, the ChiPPA focuses more on the child's ability to self-initiate and organise pretend play by itself, rather than assessing a child's social pretend play competence. Although the ChiPPA assesses some aspects of pretend play competence, unless the child initiates the involvement of the examiner it plays mostly by itself, and therefore it is mainly solitary pretend play competence that is assessed.

Another standardized and systematic observation tool for assessing children's pretend play competence is the Affect in Play Scale – Preschool (APS-P; Kaugars & Russ, 2009). The APS-P is designed for use with 4- to 6-year-old children and focuses

on affective expressions in play and cognitive components of pretend play. In a 5-minute puppet session, the child is told a short open-ended introduction story and is then encouraged to play at will. Fehr and Russ (2014) have shown that children's positive affect in play is positively associated with their prosocial behavior, as rated by their teacher. Moreover, Kaugars and Russ (2009) found positive associations between pre-schoolers' pretend play competence and their degree of integration into their peer group. Furthermore, they found negative associations with internalising problems. However, like the ChiPPA, the APS-P only assesses children's solitary pretend play.

The Test of Pretend Play (ToPP; Lewis & Boucher, 1997; formerly known as the Warwick Symbolic Play Test or WSPT) is another instrument for assessing children's pretend play competence. It differs from other instruments in that it focuses on the symbolic aspects of pretend play and offers verbal, non-verbal, structured and non-structured test versions for 1- to 6-year-olds. The ToPP is designed to assess children's ability to substitute objects and their properties as well as the ability to create scripted scenes. During the test session, an examiner encourages the child to pretend by means of play material and gives standardized prompts if needed. Clift, Stagnitti and DeMello (1998) found positive associations between the ToPP scores and age as well as language skills of 3- to 5-year-olds. However, the examiner is instructed not to interact with the child unless necessary. In sum, the above-mentioned instruments assess many important features of pretend play, including symbolism and quality of play, but they do not cover the social aspect of pretend play adequately.

The study by Smilansky (1968) was one of the first, and still one of the few, to focus on observing children's pretend play competence in a social context, thus considering the interactions between children. The Smilansky Scale for the Evaluation

of Dramatic and Socio-Dramatic Play (SSEDSP; Smilansky & Shefatya, 1990) assesses the social pretend play competence of 3- to 8-year-olds over 30 minutes of naturally occurring free play. A trained observer rates children's pretend play competence based on the occurrence of substitutions and role play as well as interactions. This instrument therefore covers different facets of social pretend play competence. Positive associations of the SSEDSP were found for example with behavioral self-regulation (Elias & Berk, 2002; Matthews, 2008; Slot, Mulder, Verhagen, & Leseman, 2017). However, the SSEDPS is a mere observation tool. Although its psychometric criteria can be judged as good (Smilansky & Shefatya, 1990), there is no standardized setting for the SSEDPS.

Another assessment tool that considers the social aspect of pretend play is the Play in Early Childhood Evaluation System (PIECES, Kelly-Vance & Ryalls, 2005). The PIECES can be used to observe children's pretend play competence for 30-45 minutes in a naturalistic but standardized play setting, in which the child is given a set of toys in the presence of its parents and an examiner. Seaborne-Borda, Kelly-Vance and Ryalls (2010) found positive associations between children's pretend play competence and their anger/frustration and attentional shifting, while no associations were found between their pretend play competence and emotion regulation. However, the authors define pretend play competence through the quantity of combined "simple pretend play" actions. This definition covers only one feature of social pretend play competence, namely the quality of sequencing the pretend play, rather than children's overall level of pretence. Furthermore, only a few items cover the social aspect of play. Moreover, the parents of the child are asked not to interact actively with the child unless it initiates social play by itself, and the setting is therefore rather unnatural for the child. Important

features like the quality of role play are only sketchily covered. Therefore, the PIECES focuses on the cognitive aspects of pretend play.

A further, recently developed measurement tool that can be used to assess children's competence is the Tools of the Play Scale (ToPS; Seeger & Holodynski, 2016). The ToPS assesses the quality of different pretend play tools such as the ability to roleplay or to substitute objects. It therefore focuses on the social-cognitive aspects of pretence. In a highly standardized play scenario test, an examiner prompts the child's use of different pretend play tools following a standardized script. Accordingly, the test considers the quality of children's pretend play as well as its social interaction component. It also prompts children to show their *highest* level of social pretend play complexity. It therefore meets our requirements for the assessment of social pretend play competence as a multidimensional construct. Good psychometric test criteria for the factorial structure and reliability of the ToPS have been confirmed (Hermann, 2017). Nevertheless, the test is conducted by an adult person and thereby the social aspect involves play interactions only between the child and an adult. Since peers are one of the most frequent play partners in early childhood and pretend play tends to occur especially while playing with peers, it is also important to consider children's pretend play competence in peer play situations.

To overcome this gap, our research group developed a new measurement tool, namely the Dyadic Pretend Play Assessment (DPPA). It assesses children's social pretend play competence in a peer play context and thereby mainly focuses on the social aspect of social pretend play. Unlike the high-standardized ToPS, where standardized prompts are set to evoke a child's use of pretend play actions and the plot is prescribed, the DPPA encourages two children to play with each other by means of a short,

standardized, open-ended instruction story and a set of partly structured play figures and material. This leads to a mainly free play situation where children have the opportunity to display their everyday play behaviour; the DPPA is therefore only half-standardized. The examiner's involvement is kept minimal and children's naturally occurring peer play can be assessed. Children's videotaped pretend play competence is evaluated afterwards by considering the developmental steps of social pretend play as defined in the current paper. The DPPA therefore assesses children's pretend play competence as a multidimensional construct. Consequently, the DPPA as well as the ToPS constitute two newly developed, promising but disparate instruments, with the ToPS focusing on social-cognitive aspects and the DPPA focusing on the social aspect of social pretend play competence. These two instruments therefore should be considered if social pretend play competence is to be assessed.

However, the DPPA and the ToPS are classic tests using adult examiners, which creates a testing atmosphere for the children. Despite the numerous advantages like the objective evaluation of children's performance, these tests may underestimate the social pretend play competence of shy and less sociable children (Crozier & Perkins, 2002). Another possibility to bypass this effect and to measure children's social pretend play competence is to observe and rate their everyday behavior in their peer group. For this purpose, a teacher's appraisal of children's social pretend play competence could be used as an additional indicator for children's pretend play competence and will be considered in the present study.

To sum up, there is a set of instruments to assess young children's pretend play competence in early childhood but none of the described pretend play measurements assesses children's social pretend play competence in a holistic, reliable and valid way.

Some of the instruments mentioned assess children's social pretend play competence only partly while others are not sufficiently validated or do not have a standardized setting. To overcome this gap, the present study examines three newly developed and most promising instruments for the assessment of the above-mentioned features of children's social pretend play competence – the ToPS, the DPPA, and a teacher report of children's social pretend play competence (RPPC) – for their psychometric properties.

Given that these three instruments differ in some aspects of assessment, their combination might yield a more holistic assessment of children's social pretend play competence. We therefore examine whether these three instruments can be used to assess and model a higher-order variable that represents children's pretend play competence (Hunter & Brewer, 2015).

Research questions

The current study aimed to evaluate, compare, and combine different methods of the assessment of social pretend play competence that vary in structure, play partner, themes and materials. We investigate factorial validity, measurement invariance, sensitivity to change, and criterion validity in terms of associations with observed social pretend play skills, social skills, and social-cognitive skills, examining the ToPS, the DPPA, and the RPPC separately. We also investigate whether combining the three assessment methods yields a higher-order pretend play competence factor. If the analyses confirm a higher-order factor, we also investigate its criterion validity. A high criterion validity is considered as given if the expected associations between social pretend play competence and children's social cognitive (ToM and emotion

understanding), social skills (sociability, cooperation, positive peer relations, emotion regulation) and language comprehension can be found.

Methods

Procedure

Eight playgroups were recruited for participation in an IRB-approved study to validate our instruments and to check the feasibility of a newly developed play intervention.

In Switzerland, playgroups are public educational institutions that are open to all children above three years. Participation is mainly funded by the parents. The aim is to give children the opportunity to play with peers under the supervision of a trained playgroup educator (Feller-Länzlinger et al., 2013). Playgroup sessions usually take place once to twice a week and last two to three hours. Children usually start in August of the year before they attend the mandatory kindergarten, and then attend the same playgroup for one year. Some children start earlier and attend the playgroup for two years.

As a first recruitment step, 42 randomly selected playgroup educators were asked to participate in the study, from which 8 playgroups confirmed. In a second step, parents with children in these playgroups were informed about the study in writing and provided written consent for their child's participation. Children were tested by trained undergraduate students pursuing an early childhood education or a psychology degree, who were sensitized to children's decision-making rights and capabilities. Additionally, children were informed about the study in an age-appropriate way (e.g. the rationale of using a video camera was explained). Children were given the opportunity to refuse or stop their individual tests at any time without any consequences. Children participated in two assessments three months apart (t1 and t2). Educators completed a questionnaire

on children's social behavior and children participated in standardized tests in individual sessions. The intervention took place in four of the eight playgroups and was similar to the intervention described in Perren, Sticca, Weiss-Hanselmann, and Burkhardt Bossi (2019). Over six consecutive weeks a play tutor visited the playgroups once a week for 30 minutes. The play tutor brought a standardized set of roleplay material, and both encouraged and guided the children in their play. Educators received financial compensation (about 8 euros per child). Children were given a small children's book as a present.

Participants

Eight playgroups from a rural region with a predominantly middle-class population participated in the study. Playgroup educators were between 34 and 52 years old (M = 43.4; SD = 6.7), trained in pedagogical work with young children, and had worked for an average of 7.9 years (SD = 5.4) as playgroup educators. All educators were Swiss. A total of 65 children (40 girls) out of 69 were given permission to participate in the study by their parents. One of the 65 children refused to participate and was not assessed. At both t1 and t2 the sample included 64 children, whereby two children left the playgroup and two new children joined it between the two measurement points. Children were between 37 and 56 months old (M = 46.4, SD = 3.8). A range of 7 to 11 children of each playgroup participated in the study, except for one playgroup where only 4 children were given permission to participate and one child refused to do so. Children's mean age ranged from 45.0 (SD = 2.6) to 48.6 (SD = 6.4) months between groups. 82% of children spoke German as family language. 18 % were learning German as a second language and had a basic knowledge of German.

Measures

Assessment of social pretend play competence

Tools of the Play Scale. The Tools of the Play Scale (ToPS) is a role play assessment for 3- to 7-year-old children, comprising a highly standardized role play with an unknown examiner accurately following a script for standardized interactions with the children. The videotaped test assesses children's social pretend play competence by assessing the quality of their use of different "tools of the play", which the authors describe as specific abilities to engage in role play. The plot is divided into two storylines: at the beginning a classic toy shop scenario is played, with the examiner playing a customer and the child playing the seller. Then a burglary happens and the examiner plays a police officer, with the child remaining the seller. The examiner sets one main prompt for each item and if necessary further prompts to evoke children's use of "the tools of the play". These tools are *substitution of objects and role-taking, substitution of play actions, substitution of role-speech* and *substitution of emotions*. These four scales are formed from the mean score of three to four items per scale (Table 1). The four further items "shopping theme", "burglary theme", "substitution of money", and "play action of combing" were excluded from the analyses both because an explorative factor analysis yielded no affiliation to one of the scales, and because some of these items were not adapted to the Swiss dialect which led to verbal misunderstandings of the prompts and therefore to a systematic bias. [Table 1 near here].

Children's videotaped use of the tools was evaluated with a standardized manual. Using this manual, children's quality of pretend play, which the examiner prompted during the ToPS, was rated for each item on a 4-point Likert-scale (0 = no

reaction despite all prompts, 1 = immature reactions with the aid of all prompts, 2 = immature reactions without the aid of all prompts or mature reactions with the aid of all prompts, 3 = mature reactions without the aid of further prompts). The four scales mentioned were used to indicate children's social pretend play quality in the subsequent latent ToPS model. The overall internal consistency of the four scales was good ($\alpha_{t1/t2}$ = .87/.81; $M_{t1/t2}$ = 10.24/13.54, $SD_{t1/t2}$ = 4.33/4.07). The reliability of the rating scales was assessed by double coding 20% of the observation cycles. The inter-rater reliability was satisfactory: intra-class correlations for all subscales were above .88 (t1) and .64 (t2).

Dyadic Pretend Play Assessment. The Dyadic Pretend Play Assessment (DPPA) assesses children's social pretend play competence in a dyadic setting while playing with a randomly selected peer. The DPPA was developed by our research group (MANUAL link). Its feasibility was checked in a small pilot study, in which small adjustments were also made to the introduction story and coding scheme. Its validity was supported and is described in the manual. In the 5-min videotaped assessment, two children are introduced to a short, standardized story by an unknown examiner. The introduction story is about the "panda family" going to have a picnic at the lake. It is played out partly by the examiner and partly by the children by means of small animal figures and unstructured materials (e.g. blue sheet). After the short, open-ended introduction story, children are invited to play on the story by themselves with the help of further play figures. For five minutes, children are totally free to act out their own ideas and play together; the assessment is therefore only half-standardized. Children's videotaped social pretend play competence is evaluated afterwards based on a standardized manual developed by our research group (MANUAL). The manual is based on the five identified features and steps of the normative development of social

pretend play competence (Bodrova et al., 2013; Smilansky & Shefatya, 1990). It consists of the categories and response options shown in Table 2. [Table 2 near here]

The individual child's behavior was rated for each minute, resulting in up to 5 cycles per child. For each category, the highest observed level of pretend play was rated and afterwards a mean of each subscale was built over the 5 cycles. Since planning was seldom observed, it was excluded from further analyses. The four remaining subscales revealed a good internal consistency of a child's social pretend play competence ($\alpha_{t1/t2}$ = .90/.85; $M_{t1/t2}$ = 2.07/2.25, $SD_{t1/t2}$ = .68/.54). As children's play behavior varied partly depending on their play partner, we assessed the DPPA two times successively with two different play partners. For each child, each subscale score was averaged out of two different partner sessions. The averaged subscale scores were than used to indicate children's social pretend play competence in the subsequent latent DPPA model. The reliability of the rating scales was assessed by double coding 20% of the observation cycles. The inter-rater reliability was high: intra-class correlations for all categories were above .94 (t1) and .80 (t2).

Teacher-reported social pretend play competence. Playgroup educators completed a short questionnaire on children's social pretend play competence, developed by our research group. The reported social pretend play competence (RPPC) scale encompasses two items on the *social quantity* and the overall *level* of children's pretend play (e.g. "how often does the child show pretend play with others", $\alpha_{t1/t2}$ = .80/.78). To ensure that all playgroup educators knew what social pretend play competence was, we provided a detailed description of the theoretical definition of low and high social pretend play competence so that the educators could better evaluate the construct. The two items were rated on a five-point Likert-type scale (e.g. 0 = never, 1= seldom, 2 =

sometimes, 3 = often, 4 = always; $M_{t1/t2}$ = 2.44/2.52, $SD_{t1/t2}$ = .91/.80) and used to indicate children's social pretend play competence in the subsequent latent RPPC model.

Tests of children's social-cognitive and language skills

Theory of Mind. ToM skills were tested using the German version of the Extended Theory of Mind Scale (EToM; Henning, Hofer, & Aschersleben, 2012; Peterson, Wellman, & Slaughter, 2012). The EToM is a well-validated test that consists of multiple tasks of increasing difficulty tapping into different developmental stages of ToM (Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006). For the study with 3- and 4-year-old children, only the first four tasks which are developmentally recommended by the authors for this age range were selected: *diverse desire, diverse beliefs, knowledge access, contents false belief.* For each task passed the children received 1 point. The mean of tasks passed was used as ToM-score for the analyses ($M_{t1/t2}$ = .48/.50, $SD_{t1/t2}$ = .29/.29, r_{t1t2} = .35, p<.05).

Emotion understanding. To test children's emotion understanding, the German Version of the Test of Emotion Comprehension was used (TEC; Janke, 2007; Pons & Harris, 2000). The TEC is a widely used test that is sensitive to individual differences in age and language (Pons, Harris, & de Rosnay, 2004), and shows good test-retest reliability (Pons & Harris, 2005). For the 3- and 4-year-old study participants, the subtests "*recognition*" and "*external cause*" (e.g. "Can you show me the sad person?"), which are developmentally recommended for this age range, were used. The mean of correct responses to the 10 items was used as emotion understanding score for the analyses ($\alpha_{t1/t2}$ = .64/.72; $M_{t1/t2}$ = .51/.61, $SD_{t1/t2}$ = .24/.25, r_{t1t2} = .73, p<.01).

Language comprehension. Children's language comprehension was tested using the German language development test for 3- to 5-year-olds (SETK 3-5; Grimm, Aktas, & Frevert, 2010). The SETK 3-5 is a widely used and standardized test for the language assessment of German-speaking children. For the present study, the first 14 items recommended for 3–4 years old of the subtest "*sentence comprehension*" were used (e.g. "Can you show me the picture, of the dog running?"; $\alpha_{t1/t2}$ = .73/.86; $M_{t1/t2}$ = .63/.72, $SD_{t1/t2}$ = .23/.24, r_{t1t2} = .75, p<.01). The mean of correct responses to these 14 items was used as sentence comprehension score for the analyses.

Teacher reports on children's social-emotional skills

Children's social-emotional skills were rated by playgroup educators. Playgroup educators completed a questionnaire on children's social behavior and emotion regulation skills (SOCOMP: Self- and Other-oriented social COMPetences, Perren, 2007; ECR: Emotion Regulation Checklist, Shields & Cicchetti, 1997). For the current study, we used the subscales *sociability*, *cooperative behavior with peers,* and *positive peer relations* from the SOCOMP and the subscale *emotion regulation* from the ECR. All items were rated on a 3-point scale (0 = not at all true, 1 = partially true, 2 = definitely true). At t2, 13 reports from two playgroups on children's social-emotional skills could not be acquired because they were not provided by the educator. The *sociability* subscale, covering propensity to participate in social interactions, consists of four items (e.g. "Converses with peers easily"; $\alpha_{t1/t2}$ = .77/.79; $M_{t1/t2}$ = 1.47/1.61, $SD_{t1/t2}$ = .52/.50, r_{t1t2} =.78, p<.01). The *cooperation with peers* subscale consists of three items (e.g. "Compromises in conflicts with peers"; $\alpha_{t1/t2}$ = .62/.77; t1: $M_{t1/t2}$ = 1.38/1.43, $SD_{t1/t2}$ = .50/.42, r_{t1t2} = .46, p<.01). The *positive peer relations* subscale consists of five items (e.g. "Has a lot of friends"). Due to a low internal consistency of $a_{t1/t2}$ = .52/.77,

one item was excluded from the scale, which gives a Cronbach's alpha of $\alpha_{t1/t2}$ = .68/.84 ($M_{t1/t2}$ = 1.57/1.70, $SD_{t1/t2}$ = .46/.50, r_{t1t2} = .67, p<.01). The *emotion regulation* subscale consists of eight items (e.g. "Can say when s/he is feeling sad, angry or mad, fearful or afraid"). Because of the low Cronbach's alpha at t1 ($a_{t1/t2}$ = .59/.67), we excluded two items (items 2 and 6) of the emotion regulation subscale, resulting in an acceptable Cronbach's alpha of $\alpha_{t1/t2}$ = .67/.71 *(*$M_{t1/t2}$ = 1.63/1.70, $SD_{t1/t2}$ = .31/.32, r_{t1t2} = .70, p<.01).

Statistical analyses

AMOS 25 was used for all analyses. The analyses were made using a full information maximum likelihood estimation. First, social pretend play competence, measured twice by the different assessment tools, was modelled as a latent variable for each tool, based on the theoretical and empirical assumptions of the factor structures made by the instrument's authors. The models were then tested for their factorial validity and measurement invariance by confirmatory factor analyses. Where indicated, small adjustments were made to the models to achieve at least partial scalar measurement invariance. The test-retest reliability was examined, as well as sensitivity to change of each instrument. Second, all individual measurement factors were combined to form a higher-order latent factor in order to test whether we can identify an underlying factor of social pretend play competence. Third, to establish and compare the criterion validity of the different measures, bivariate associations between the individual social pretend play competence measures and children's social-cognitive skills (ToM skills, emotion understanding, language comprehension), social-emotional skills (sociability, cooperation, peer relations, and emotion regulation), age and sex were calculated and compared.

We did not anticipate changes in the content-related meaning of social pretend play competence as a psychological construct due to the intervention and therefore expected to find the same factor structure for control and intervention groups at post-test. This assumption was tested by multi-group comparisons between intervention and control groups at t2. The measurement models as well as the covariances between social pretend play measures and the criterion variables were strongly invariant between control and intervention group at t2 for the ToPS ($\Delta \chi^2 = 13.54$, $\Delta df = 14$, $p = .49$), the DPPA ($\Delta \chi^2 = 17.51$, $\Delta df = 16$, $p = .35$), as well as the RPPQ ($\Delta \chi^2 = 18.45$, $\Delta df = 11$, $p = .08$). Moreover, we considered the impact of intervention on our test-retest and sensitivity to change analysis by controlling for the experimental condition. Analyses revealed no main intervention effects.

Results

Factorial validity and reliability of the single social pretend play competence measures

The factorial validity of the three different social pretend play competence measurements was analyzed separately for each instrument. Measurement invariance, test-retest reliability, and sensitivity to change of the different instruments were examined, including t1 and t2 indicators in each model. Correlated uniqueness between t1 and t2 scores of the same variables was freely estimated.

Tools of the Play Scale (ToPS). For the ToPS, the four indicators *substitution of objects, substitution of play actions, substitution of speech* and *substitution of emotion* were used to indicate social pretend play competence. The model fit well to the data and the analyses yielded weak measurement invariance for the ToPS. Modification indices were then used to examine which intercepts had to be released to achieve a satisfactory model

fit for partial scalar invariance (Yoon & Kim, 2014). The "substitution of emotion" subscale was found to have no invariant item intercept. Therefore, the intercept of "substitution of emotion" subscale was released to achieve partial scalar invariance (see Table 3). The bivariate correlation between the t1 and t2 score showed a good test-retest reliability for the ToPS ($r(47) = .72$, $p < .01$), controlled for intervention. Next, a latent change model was calculated to test the measurement instruments for their sensitivity to change. Social pretend play competence assessed through the ToPS showed a significant latent change from t1 to t2 ($M_{t1/t2} = 10.24/13.54$, $SD_{t1/t2} = 4.33/4.07$; $\Delta_{t1/t2} = 2.90$, $p<.01$) and a significant variance of the latent change showing individual differences of change scores between children ($\sigma_{ch}^2 = 7.59$, $p < .01$), controlled for intervention.

Dyadic Pretend Play Assessment (DPPA). For the DPPA model, the following four indicators were used to build the latent factor: *decentration, decontextualization, role taking* and *sequencing*. The resulting latent DPPA variable showed a good fit to the data and weak measurement invariance. Modification indices were then used to examine which intercepts had to be released to achieve a satisfactory model fit for partial scalar invariance. The decontextualization subscale was found to have no invariant item intercept. The intercept of decontextualization was therefore released to achieve partial scalar invariance (see Table 1

Overview of Scales and Items of the Tools of *the Play Scale*

Substitution Scale	Item		Tools	$\alpha t1_{/t2}$
Objects and Role-Taking				
	P1-2a	Substituting blocks as toys	OS	
	P1-3	Recommending suitable gift for opposite gender	RO	
	P1-3a	Offering further gift	RO	
				.58/.57
Play Actions				
	P1-6	Wrapping gift	PA	
	P1-7	Substituting block as chocolate bar	PA	
	P1-8	Arranging the payment	PA	

				.62/.49
Role-Speech	P1-5	Substituting block as book and reading from it	RL	
	P2-4	Calling police concerning the burglary	RL	
	P2-6	Reporting the police's answer to customer	RL	
				.80/.78
Emotions	P2-2	Shock about the burglary	EE	
	P2-3	Anger towards the burglar	EE	
	P2-5	Relief about the captured burglar	EE	
	P2-7	Sympathy for the sad customer	EE	
				78/.79

Note. P1= storyline 1, P2= storyline 2, RO = substitution of role, OS= substitution of object, PA= substitution of play action, RL= substitution of role language, EE = substitution of emotions, $\alpha t1_{/t2}$ = cronbach's α of scale at t1/t2.

Table 2

Simplified Representation of the Coding Scheme for DPPA Ratings

Category	Level/Value	Description
Decentration	1	No decentration
	2	Self-related pretend play action
	3	Object-related pretend play action
	4	Other-related pretend play action
Decontextualization	1	No decontextualization
	2	Object imitation
	3	Object substitution
	4	Fantasy transformation
Role-taking	1	No role-taking behavior
	2	Role-taking without role-conforming behavior
	3	Role-taking with some role-conforming behavior
	4	Sustained role-taking and role-conformity

Planning	1	No planning
	2	Planned single play action without performance
	3	Planned and performed single play actions
	4	Several connected planned and performed play actions
Sequencing	1	No sequencing
	2	Limited script elements are performed
	3	Several related actions are performed
	4	Flexible application of a script to the play

Note. DPPA = Dyadic Pretend Play Assessment.

). Additionally, the model showed a good test-retest reliability for the DPPA ($r(53)$ = .72, $p<.01$), controlled for intervention. A significant change in pretend play competence was found for the DPPA ($M_{t1/t2}$ = 2.07/2.25, $SD_{t1/t2}$ = .68/.54, $\Delta_{t1/t2}$ = .18, $p<.05$). Additionally, the variance of the change was significant (σ_{ch}^2 = .19, $p<.01$), controlled for intervention.

Teacher-reported social pretend play competence (RPPC). The latent pretend play competence from the educator's questionnaire was formed out of the two items *social quantity* and *quality of play*. To model the reported social pretend play competence as a latent construct, both indicators' loadings had to be set to 1 in order to achieve local identification. As a result of the equation of the two indicators of the RPPC, the model of the RPPC was only tested for scalar invariance, which was confirmed by the model fit (see Table 3). Further, the model of the RPPC showed a good test-retest reliability ($r(59)$ = .71, $p<.01$), controlled for intervention. No significant change in pretend play competence was found for the RPPC ($M_{t1/t2}$ = 2.44/2.52, $SD_{t1/t2}$ = .91/.80, $\Delta_{t1/t2}$ = .08; p = .54), but the variance of the change was significant (σ_{ch}^2 = 0.24, $p < .01$), controlled for intervention. Therefore, the RPPC seems to be sensitive to differences in changes between children. [Table 3 near here]

Establishing a second-order factor of social pretend play competence

The three first-order latent variables described above were modelled as latent indicators of a latent second-order factor of children's social pretend play competence. The model for the first measurement point (t1) is shown in Figure 1. [Figure 1 near here]

The model fit confirmed an acceptable fit of the data (χ^2 = 50.68, df = 33, CFI = .95, RMSEA= .09, 90% CI: [0.03, 0.14]). The different standardized factor loadings were .82 for the ToPS, .54 for the DPPA, and .53 for the RPPC. The model pertaining to

the second measurement point was modelled in the same way. The good fit of the model to the data was also confirmed (χ^2 = 38.31, df = 33, CFI = .98, RMSEA = .05, 90% CI: [.00, .11]).

Criterion validity

The results of the bivariate correlation with different social-cognitive and social skills variables as well as with age and sex are shown in Table 4. [Table 4 near here]

Tools of the Play Scale. The ToPS showed positive associations with a child's ToM (only at t2), emotion understanding, language comprehension, sociability, positive peer relations and emotion regulation. A marginally significant positive association was found with age at t1. No associations were found between the ToPS score and cooperation or sex.

Dyadic Pretend Play Assessment. The DPPA was found to be positively associated with children's emotion understanding (only at t2), sociability (only at t1), as well as marginally with language comprehension (only at t1), their positive peer relations (only at t1), and age. No associations were found between the DPPA and ToM, cooperation, emotion regulation, or sex.

Teacher-reported pretend play competence. The RPPC showed positive associations with children's ToM, emotion understanding, language comprehension, sociability, positive peer relations, and emotion regulation. No associations were found between the RPPC and children's cooperation, age or sex.

Pretend play competence (higher-order factor). The combination of the three assessments as latent pretend play competence showed moderate positive correlations with the tested social-cognitive skills and (very) strong positive correlations with all

social behavior scales except cooperation. In addition, no associations were found with age or sex.

Discussion

The present study examined the psychometric properties of three social pretend play assessment methods: ToPS, DPPA, and RPPC. The results showed that all the instruments are appropriate for the assessment of children's social pretend play competence and are positively correlated with each other, but also that they partly cover different facets of social pretend play competence.

Tools of the Play Scales. The ToPS appear to cover children's social pretend play competence adequately. A high criterion validity for the ToPS was found for ToM, emotion understanding and language comprehension, as well as for sociability, positive peer relations and emotion regulation. This is in line with the theoretical assumption that children's social pretend play competence is positively related to their social-cognitive and social development (Piaget, 1962; Vygotsky, 1978). Furthermore, the ToPS showed stronger correlations with children's social-cognitive skills. This confirms the social-cognitive focus of the ToPS. No association was found between cooperation and sex, which will be discussed below. Furthermore, the ToPS was sensitive to change. However, our examiners reported that during the assessments at t2 children remembered some of the tested items, and so a learning effect cannot be ruled out. To prevent memory effects, we recommend developing a parallel version of the ToPS.

A strength of the ToPS is that its demanding character means it assesses children's highest level of social pretend play competence. By prompting their use of the different tools during the test, children are encouraged to show their highest competence level, in other words, the most mature pretend play of which they are

capable. However, from a practitioner's point of view, the ToPS is not easy to implement, as the prompts and their order must be learned by heart and set standardized. Administration of the ToPS is also challenging, because children often deviate from the intended plot. The examiner has to keep the children motivated by mentioning their ideas but must also follow the script and set the standardized prompts. The examiners thus need to be trained well. On the other hand, the ToPS itself takes 20 minutes to conduct, which is appropriate for young children. Another point that should be considered is that the ToPS is set up by an unknown examiner, so that the testing situation could cause scores to be underestimated (Crozier & Perkins, 2002). Furthermore, the ToPS requires a basic level of speech, which is beneficial for testing children with different speech levels; on the other hand, children can only be tested if they already have a certain basic level of speech.

To sum up, the ToPS seems to be a reliable and valid instrument for assessing children's social pretend play competence, and especially for testing their highest level of social pretend play competence. The ToPS is therefore particularly useful to assess children's developmental status. However, in comparison with the DPPA, the ToPS does not cover the social aspect of children's *peer* pretend play during the assessment, which was slightly indicated by the weaker correlations with sociability and positive peer relations at t1. In conclusion, the ToPS covers most of the features of social pretend play competence and can be used to assess children's highest level of social pretend play competence.

Dyadic Pretend Play Assessment. The DPPA showed good factorial validity and partial scalar measurement invariance as well as moderate test-retest reliability and sensitivity to change. However, it should be noted that the DPPA is always carried out with

another peer, and therefore relationship effects cannot be ruled out. As children's play partners changed within and between t1 and t2, the test-retest reliability is still judged to be good. Furthermore, the overall change of DPPA scores could indicate that the children did reach the next developmental step of social pretend play quality as measured by the DPPA. On the other hand, it could be that children got used to the testing situation and the play material and were therefore less shy and/or more able to show higher pretend play competences in the free play situation.

Regarding the construct validity, we found positive associations with children's emotion understanding, language comprehension, sociability, and positive peer relations. These associations match the peer context in which the assessment took place and confirm the social nature of pretend play that we assessed through the DPPA. However, we expected the DPPA to correlate with more of the outcome variables to confirm its congruent validity. Furthermore, in contrast to the ToPS and RPPC, we did not find the same association scheme between t1 and t2. As children were not able to choose the peer they played with, this could be a crucial point in explaining the missing or inconsistent associations found with the DPPA. The social pretend play competence of some children could have been over- or underestimated because of their relationship with their randomly selected play partner. We tried to overcome this gap by averaging a child's play with two different children. For further investigations, we would recommend assessing even more sessions to obtain a more valid score. Additionally, because of the setting, children do not have to show their highest level of pretend play competence, in contrast to the ToPS. This could be another explanation for the low congruent validity of the DPPA. On the other hand, the setting of the DPPA is closer to children's everyday life and play, and therefore assesses children's everyday

performance. The peer context is thus both the DPPA's strength and its limitation. Another benefit of the DPPA is that it can be used to assess children with very poor language skills. Furthermore, it takes only 5 minutes and is therefore also feasible for practitioners.

To conclude, the DPPA provides an adequately valid assessment of children's social pretend play quality, but it seems to be influenced by relationship effects. It can therefore be used to assess the role of peers during dyadic pretend play. We would recommend using it alone if an investigation focuses on children's quality of social pretend play in everyday life, and three to four DPPA sessions with different play partners are averaged in order to minimise relationship effects.

Teacher-Reported Social Pretend Play Competence. The RPPC scale showed good factorial validity. For a further assessment, a third item (or even more) should be added to the questionnaire to meet the requirements for the statistical modulation of a just-identified latent variable. Positive associations were found with ToM, emotion understanding, language comprehension, sociability, positive peer relations, and emotion regulation. Moreover, high correlations with the teacher-reported social skills are noticeable. The high correlations seem to include a shared method variance. However, the non-existent association with the cooperation scale shows that it is not solely explainable by a halo effect, where every item of a questionnaire would be rated higher or lower based on a general impression of the child. Therefore, it can be assumed that there are positive associations, albeit not that strong. The results show that the RPPC seems to provide a broad assessment of children's social pretend play competence.

In conclusion, the RPPC's assessment of children's social pretend play competence is reliable and valid, and moreover, it is very economical to set up, especially in comparison with the ToPS and the DPPA. However, we cannot conclude that a questionnaire is the best method for assessing children's social pretend play competence in general. The educators in our study were unusually able to complete this questionnaire as they have a pedagogical background and witness the children most of the time during play. It is therefore not that suitable for all aims. It also contains only two items and is therefore quite undifferentiated. However, like the DPPA, the RPPC does cover children's everyday-life pretend play.

Social Pretend Play Competence (higher-order factor). Combining the three assessments mentioned revealed an underlying latent social pretend play competence score. Interestingly, the ToPS and the RPPC seem to represent the highest amount of underlying social pretend play competence. Following the positive association of the DPPA with children's sociability and positive peer relationships, the added value of the DPPA to the social pretend play competence model seems to be the peer interaction feature as proposed. The latent higher-order social pretend play competences showed the highest criterion validity. The combination of the three assessments seem to cover the different facets of children's social pretend play competence in a broadly, reliable, valid and holistic way. As long as no single instrument with such good criteria is available, we would recommend using all three of these instruments to assess children's social pretend play competence broadly, as competence and performance in adult and peer interactions. However, if the cost benefit is considered, the ToPS and the RPPC seem to achieve equally good psychometric criteria. We would therefore recommend using the ToPS to assess children's social pretend play competence, especially

diagnostically, where single tests with individual children are more practicable. The RPPC is recommended if the examiner has a pedagogical background and has experience with children in their everyday play with other children, and therefore is more suitable for the assessment of young children's pretend play competences in early childcare settings. The DPPA can be used to assess children's social pretend play performance and the role of peers in peer play settings in a Vygotskian tradition. Its simple set up makes it usable in many research and diagnostic contexts.

Associations with cooperation, sex and age. It is noteworthy that no associations with children's social pretend play competence were found for either cooperation or children's sex in the present study. The non-existent association with cooperation was unexpected but is still not surprising: Previous studies found mixed results for the relation of pretend play and cooperation (Li, Hestenes, & Wang, 2016; Lillard et al., 2013; Swindells & Stagnitti, 2006). Hence, there is no clear evidence of a positive association. Nevertheless, none of the three independent measures of children's social pretend play competence showed a positive relation to cooperation in the present study. This can be interpreted as a strong indication that there is in fact no association with children's cooperation with peers, as measured in the present study. This could mean that cooperative behavior is an important prerequisite for social pretending (Bodrova et al., 2013), but not for the *quality* of the play. The non-existent relation with sex could mean that for children aged 3-4, sex differences do not yet influence their social pretend play competence. Previous studies have shown inconsistent results concerning sex differences in pretend play (Göncü, Patt, & Kouba, 2002). Göncü et al. (2002) assume that the inconsistent findings can be traced back to the different assessments of pretend play used in the studies. The present study showed that there seem to be no sex

differences in the *performance* or *competence* of 3- to 4-year olds' social pretend play. Further, only minor age effects were found with older children scoring higher on the ToPS and DPPA, which is in line with the normative pattern of social pretend play (Howes & Matheson, 1992). The weak associations could be due to the small age range in the group. However, children's social pretend play competence was positively associated with other developmental variables, which shows that it is associated with children's developmental age.

Strengths and limitations of the study

The present study has many strengths, especially its practical implications for further research on children's pretend play, but also some limitations that need to be addressed in future studies. It is indisputable that there is urgent need for a well-validated assessment of children's social pretend play competence (Bergen, 2013; Lillard et al., 2013). The present study is the first that examines and compares three different assessment methods within one sample. It therefore makes an important contribution to the further investigation of young children's pretend play. Another strength of the present study is its design with two measurement points, which enabled the examination of the measurement invariance, its test-retest reliability, as well as the sensitivity to change of the assessments.

However, the study has a few constraints in its implementation. First of all, the sample size is at limit for the statistical analysis made. At the same time, it is noticeable that the second-order model fitted well enough with such a small sample, which in turn speaks for the underlying latent construct of children's social pretend play competence. Another limitation is the intervention between the two measurement points, which could have biased our results. However, we controlled for the intervention between

measurement points and for measurement invariance between intervention and control group at post-test, which enabled us to limit this bias. The analyses revealed no main intervention effects. Additionally, some of the reliabilities of the ToPS Scales were at limit for statistical analyses, which should be kept in mind when interpreting results. However, as we wanted to rebuild the factor structure as proposed by the authors of the ToPS, we accepted the small coefficients. Nearly 21% of the second measurement data from the teacher-reported pretend play competence were missing, which is a further limitation to consider when interpreting the data. Moreover, there seems to be a method effect within the same measurement methods as the testing scores of social pretend play competence correlated more highly with the tested outcomes, and the reported social pretend play competence correlated very highly with the reported outcomes.

Implications for practice and further research

The present study offers important insights for the assessment of children's social pretend play competence. To date, the field of pretend play research has had no common, reliable, and valid instruments for assessing children's social pretend play competence in a standardized setting. With the recommendation of this study, new standards could be set up, so that future studies could use the recommended assessments. Results of different studies could be compared and more information about associations with children's pretend play and their causality gathered to shed light onto pretend play. The study showed that a role play test and teacher reports are adequate methods for assessing children's social pretend play competence. The instruments could now be used in further studies, for example to evaluate pretend play interventions, or to investigate the role of different play partners in social pretend play quality during peer play. So far, research has been able to show that promoting children's pretend play

through an active support fosters children's social pretend play quality immediately (Perren, Sticca, Weiss-Hanselmann, & Burkhardt Bossi, 2019). The instruments discussed could now be used to investigate whether this promotion is long-lasting. However, they should also be further developed, based on our findings.

Declaration of Interest Statement

No potential conflict of interest

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable requests.

References

Bergen, D. (2013). Does pretend play matter? Searching for evidence: Comment on Lillard et al. (2013). *Psychological Bulletin*, *139*(1), 45–48. https://doi.org/10.1037/a0030246

Bodrova, E., Germeroth, C., & Leong, D. J. (2013). Play and Self-Regulation: Lessons from Vygotsky. *American Journal of Play*, *6*(1), 111–123.

Bodrova, E., & Leong, D. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed). Upper Saddle River, N.J: Pearson/Merrill Prentice Hall.

Clift, S., Stagnitti, K., & DeMello, L. (1998). A validational study of the test of pretend play using correlational and classificational analyses. *Child Language Teaching and Therapy*, *14*(2), 199–209.

Crozier, W. R., & Perkins, P. (2002). Shyness as a Factor when Assessing Children. *Educational Psychology in Practice*, *18*(3), 239–244. https://doi.org/10.1080/0266736022000010267

Elias, C. L., & Berk, L. E. (2002). Self-regulation in young children: Is there a role for sociodramatic play? *Early Childhood Research Quarterly*, *17*(2), 216–238. https://doi.org/10.1016/S0885-2006(02)00146-1

Fehr, K. K., & Russ, S. W. (2014). Assessment of Pretend Play in Preschool-Aged Children: Validation and Factor Analysis of the Affect in Play Scale–Preschool Version. *Journal of Personality Assessment*, *96*(3), 350–357. https://doi.org/10.1080/00223891.2013.838171

Fein, G. G. (1981). Pretend Play in Childhood: An Integrative Review. *Child Development*, *52*(4), 1985–1118.

Göncü, A., Patt, M. B., & Kouba, E. (2002). Understanding Young Children's Pretend Play In Context. In P. K. Smith & C. H. Hart (Hrsg.), *Blackwell handbook of childhood social development* (S. 418–437). Malden MA: Blackwell Publishing.

Grimm, H., Aktas, M., & Frevert, S. (2010). *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen* (2. Auflage). Göttingen: Hogrefe.

Henning, A., Hofer, T., & Aschersleben, G. (2012). *Erweiterte „Theory of Mind"-Skala für 3 bis 11-jährige Kinder. Authorised German translation of the Extended Theory-of-Mind Scale by Peterson, Wellman & Slaughter, 2012.* Saarland University.

Hermann, S. (2017). *Evaluation of a Fairytale-Based Pretend-Play Intervention to Foster Socioemotional Competences in Preschool Children* (Inaugural dissertation). Westfälischen Wilhelms-Universität Münster, Universitäts-und Landesbibliothek Münster.

Howes, C., & Matheson, C. C. (1992). Sequences in the development of competent play with peers: Social and social pretend play. *Developmental Psychology*, *28*(5), 961–974. https://doi.org/10.1037/0012-1649.28.5.961

Hunter, A., & Brewer, J. D. (2015). Designing Multimethod Research. In S. N. Hesse-Biber & R. B. Johnson (Hrsg.), *The Oxford handbook of multimethod and mixed methods research inquiry*. https://doi.org/10.1093/oxfordhb/9780199933624.013.13

Janke, B. (2006). *Skala zur Erfassung des Emotionswissens für 3- bis 10-jährige Kinder.* Heidelberg: Pädagogische Hochschule.

Kaugars, A. S., & Russ, S. W. (2009). Assessing Preschool Children's Pretend Play: Preliminary Validation of the Affect in Play Scale-Preschool Version. *Early Education and Development*, *20*(5), 733–755. https://doi.org/10.1080/10409280802545388

Kelly-Vance, L., & Ryalls, B. O. (2005). A Systematic, Reliable Approach to Play Assessment in Preschoolers. *School Psychology International*, *26*(4), 398–412. https://doi.org/10.1177/0143034305059017

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von "Theory of Mind"-Aufgaben. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *38*(4), 186–195. https://doi.org/10.1026/0049-8637.38.4.186

Lewis, V., & Boucher, J. (1997). *The Test of Pretend Play: ToPP*. Psychological Corporation.

Li, J., Hestenes, L. L., & Wang, Y. C. (2016). Links Between Preschool Children's Social Skills and Observed Pretend Play in Outdoor Childcare Environments. *Early Childhood Education Journal*, *44*(1), 61–68. https://doi.org/10.1007/s10643-014-0673-2

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, *139*(1), 1–34. https://doi.org/10.1037/a0029321

Lillard, A. S., Pinkham, A., M., & Smith, E. (2011). Pretend Play and Cognitive Development. In *Blackwell handbooks of developmental psychology*. *The Wiley-*

Blackwell handbook of childhood cognitive development (2nd ed). Malden, MA: Wiley-Blackwell.

Matthews, S. B. (2008). *The relationship among self-regulation, sociodramatic play, and preschoolers' readiness for kindergarten* (Inaugural dissertation). Northeastern University, Boston: Massachusetts.

Overton, W. F., & Jackson, J. P. (1973). The Representation of Imagined Objects in Action Sequences: A Developmental Study. *Child Development*, *44*(2), 309–314. https://doi.org/10.2307/1128052

Perren, S. (2007). *SOCOMP (Self- and Other-oriented social Competences). Ein Fragebogen zur Erfassung von selbst- und fremdorientierten verhaltensbezogenen sozialen Kompetenzen. Manual*. Zürich: Jacobs Center for Productive Youth Development, University of Zurich.

Perren, S., Sticca, F., Weiss-Hanselmann, B., & Burkhardt Bossi, C. (2019). Let us play together! Can play tutoring stimulate children's social pretend play level? *Journal of Early Childhood Research*, *17*(3), 205–219. https://doi.org/10.1177/1476718X19849248

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome. *Child Development*, *83*(2), 469–485. https://doi.org/10.1111/j.1467-8624.2011.01728.x

Piaget, J. (1962). *Play, Dreams & Imitation in Childhood*. Abgerufen von http://www.amazon.com/Dreams-Imitation-Childhood-Library-Paperback/dp/0393001717

Pons, F., & Harris, P. (2000). *Test of emotion comprehension: TEC*. University of Oxford.

Pons, F., & Harris, P. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition & Emotion*, *19*(8), 1158–1174. https://doi.org/10.1080/02699930500282108

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152. https://doi.org/10.1080/17405620344000022

Seaborne-Borda, L., Kelly-Vance, L., & Ryalls, B. (2010). *Preschoolers' Emotion Regulation and Negative Affect: Links to Levels of Play?* (S. 24). Omaha: University of Nebraska.

Seeger, D., & Holodynski, M. (2016). *Diagnoseverfahren zur Rollenspielkompetenz von Vorschulkindern. Manual zur Durchführung und Auswertung. Unveröffentlichtes Manuskript*. Münster: Institut für Psychologie in Bildung und Erziehung der Universität Münster.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology*, *33*(6), 906–916. https://doi.org/10.1037/0012-1649.33.6.906

Slot, P. L., Mulder, H., Verhagen, J., & Leseman, P. P. M. (2017). Preschoolers' cognitive and emotional self-regulation in pretend play: Relations with executive functions and quality of play. *Infant and Child Development*, *26*(6), e2038. https://doi.org/10.1002/icd.2038

Smilansky, S. (1968). *The Effects of Sociodramatic Play on Disadvantaged Preschool Children*. New York: Wiley.

Smilansky, S., & Shefatya, L. (1990). *Facilitating play: A medium for promoting cognitive, socio-emotional and academic development in young children*. Psychological & Educational Publications.

Smith, P. K. (2010). *Children and Play: Understanding Children's Worlds*. Malden, MA: Wiley-Blackwell.

Stagnitti, K., & Unsworth, C. (2004). The Test-Retest Reliability of the Child-Initiated Pretend Play Assessment. *American Journal of Occupational Therapy*, *58*(1), 93–99. https://doi.org/10.5014/ajot.58.1.93

Stagnitti, K., Unsworth, C., & Rodger, S. (2000). Development of an assessment to identify play behaviours that discriminate between the play of typical preschoolers and preschoolers with pre-academic problems. *Canadian Journal of Occupational Therapy*, *67*(5), 291–303.

Swindells, D., & Stagnitti, K. (2006). Pretend play and parents' view of social competence: The construct validity of the Child-Initiated Pretend Play Assessment. *Australian Occupational Therapy Journal*, *53*(4), 314–324. https://doi.org/10.1111/j.1440-1630.2006.00592.x

Taylor, M., Carlson, S. M., Maring, B. L., Gerow, L., & Charley, C. M. (2004). The Characteristics and Correlates of Fantasy in School-Age Children: Imaginary Companions, Impersonation, and Social Understanding. *Developmental Psychology*, *40*(6), 1173–1187. https://doi.org/10.1037/0012-1649.40.6.1173

Vygotsky, L. S. (1978). *Mind in Society. Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Yoon, M., & Kim, E. S. (2014). A comparison of sequential and nonsequential specification searches in testing factorial invariance. *Behavior Research Methods*, *46*(4), 1199–1206. https://doi.org/10.3758/s13428-013-0430-2

Figures and Tables

Table 1

Overview of Scales and Items of the Tools of the Play Scale

Substitution Scale	Item		Tools	$\alpha_{t1/t2}$
Objects and Role-Taking				
	P1-2a	Substituting blocks as toys	OS	
	P1-3	Recommending suitable gift for opposite gender	RO	
	P1-3a	Offering further gift	RO	
				.58/.57
Play Actions				
	P1-6	Wrapping gift	PA	
	P1-7	Substituting block as chocolate bar	PA	
	P1-8	Arranging the payment	PA	
				.62/.49
Role-Speech	P1-5	Substituting block as book and reading from it	RL	
	P2-4	Calling police concerning the burglary	RL	
	P2-6	Reporting the police's answer to customer	RL	
				.80/.78
Emotions	P2-2	Shock about the burglary	EE	
	P2-3	Anger towards the burglar	EE	
	P2-5	Relief about the captured burglar	EE	
	P2-7	Sympathy for the sad customer	EE	
				78/.79

Note. P1= storyline 1, P2= storyline 2, RO = substitution of role, OS= substitution of object, PA= substitution of play action, RL= substitution of role language, EE = substitution of emotions, $\alpha_{t1/t2}$ = cronbach's α of scale at t1/t2.

Table 2

Simplified Representation of the Coding Scheme for DPPA Ratings

Category	Level/Value	Description
Decentration	1	No decentration
	2	Self-related pretend play action
	3	Object-related pretend play action
	4	Other-related pretend play action
Decontextualization	1	No decontextualization
	2	Object imitation
	3	Object substitution
	4	Fantasy transformation
Role-taking	1	No role-taking behavior
	2	Role-taking without role-conforming behavior
	3	Role-taking with some role-conforming behavior
	4	Sustained role-taking and role-conformity
Planning	1	No planning
	2	Planned single play action without performance
	3	Planned and performed single play actions
	4	Several connected planned and performed play actions
Sequencing	1	No sequencing
	2	Limited script elements are performed
	3	Several related actions are performed
	4	Flexible application of a script to the play

Note. DPPA = Dyadic Pretend Play Assessment.

Table 3

Model Fit Indices for the Three Levels of Measurement Invariance

	Model	χ^2	df	p (χ^2)	CFI	RMSEA [90 % CI]
ToPS	Configural	19.05	18	.39	.99	.03 [.00, .12]
	Metric[1]	22.88	21	.35	.99	.04 [.00, .11]
	Partial scalar[2]	26.83	23	.26	.98	.05 [.00, .12]
DPPA	Configural	25.04	18	.12	.98	.08 [.00, .14]
	Metric[1]	27.71	21	.15	.98	.07 [.00, .13]
	Partial scalar[2]	30.46	23	.14	.98	.07 [.00, .13]
RPPC	Configural	3.82	2	.15	.98	.12 [.00, .30]
	Metric[1]	3.82	2	.15	.98	.12 [.00, .30]
	Scalar[2]	5.52	3	.14	.98	.12 [.00, .26]

Note. [1] Factor loadings set equal among time points; [2] Factor loadings and item intercepts (partly) set equal among time points. CI = Confidence Interval; ToPS = Tools of the Play Scale; DPPA = Dyadic Pretend Play Assessment; RPPC = Reported Pretend Play Competence.

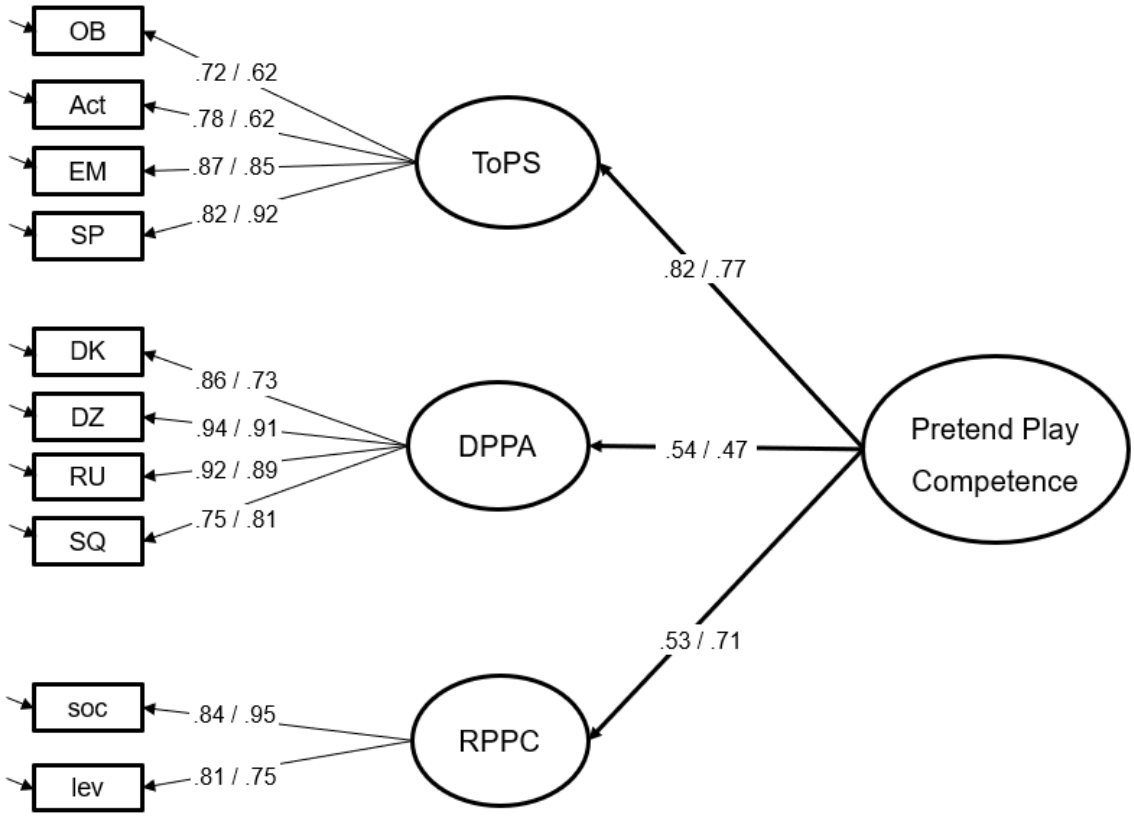

Figure 1. Second-order model of children's social pretend play competence (t1 / t2 data). ToPS = Tools of the Play Scale; DPPA = Dyadic Pretend Play Assessment; RPPC = Reported Pretend Play Competence.

Table 4

Bivariate correlations of social pretend play measures (latent) and criterion variables (t1 & t2)

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	ToPS	1	.43*	.45*	.82*	.14	.53**	.59***	-.01	.34*	.21	.38*	*.25*†	.06
2	DPPA	.36*	1	*.29*†	.54*	-.01	.23	*.24†*	-.05	.29*	*.27†*	.00	*.24*†	-.07
3	RPPC	.51**	.31*	1	.53*	.37*	.31*	.29*	.03	.74***	.83***	.60***	.06	.22
4	2nd-order factor	.77*	.47*	.71*	1	.37*	.41*	.40*	.02	.81***	.87***	.63***	.13	.20
5	ToM	.35*	-.03	.31*	.37*	1	.47**	.52***	-.03	.09	.12	*.23*†	.31*	.09
6	TEC	.49**	.33*	.45**	.58**	.36*	1	.70***	.03	.14	.14	.31**	.36*	*.22*†
7	SETK	.33*	.21	.50**	.55**	.40**	.62***	1	.10	.06	.10	.30**	.15	.04
8	Cooperation	.10	-.10	.05	.06	.51**	.32*	*.19*†	1	.00	.14*	.26*	-.07	.27*
9	Sociability	.37*	-.05	.60***	.65***	*.24*†	.19	*.24*†	.17	1	.72***	.62***	-.10	.13
10	Positive peers	.33*	-.10	.68***	.70***	.38**	*.27*†	.43**	.29*	.80***	1	.58*	-.08	.19
11	Emotion regulation	.43**	-.15	.47**	.54**	*.25*†	.31*	.44**	.33*	.68***	.60***	1	.06	.12
12	Age	.19	.18	.07	.13	.22	*.23*†	-.04	-.08	-.14	-.13	-.10	1	.07
13	Female	.20	-.18	.21	.23	.28*	.18	.05	.30*	.34*	.42**	.13	.10	1

Note. Correlation coefficients above the diagonal are cross-sectional correlations of t1 variables. Correlation coefficients under the diagonal show correlations of t2 variables. ToPS = Tools of the Play Scale; DPPA = Dyadic Pretend Play Assessment; RPPC = Reported Pretend Play Competence; ToM = Theory of Mind; TEC = Test of Emotion Comprehension; SETK = Language Comprehension; Female = sex with "0 = male" and "1 = female". †p<.10; *p<.05; **p<.01; ***p<.001.