The Test–Retest Reliability of the Child-Initiated Pretend Play Assessment

Karen Stagnitti,
Carolyn Unsworth

OBJECTIVE. The purpose of this study is to establish the test–retest reliability of the Child-Initiated Pretend Play Assessment (ChIPPA) (Stagnitti, 2002a; Stagnitti, Unsworth, & Rodger, 2000).

METHOD. The first author rated 38 preschool children ages 4 and 5 years (4 with developmental delay and 34 typically developing) on the ChIPPA. The ChIPPA employs conventional play materials and unstructured play materials to assess three qualities of a child's play ability: elaborateness of play action, ability to substitute objects during play, and the child's need to imitate the modelled actions of the examiner. The ChIPPA was administered twice, at a 2-week interval, to each participant.

RESULTS. Test–retest intraclass correlation coefficients (ICCs) (Type 2,1) calculated for each of the three elaborate play measures ranged from .73 to .84. A test–retest ICC of .56 was obtained for object substitution with unstructured play materials. The test–retest ICC obtained for the combined score for unstructured and conventional play materials was .57. Percentage agreement figures ranging from 63.2% to 84.2% were obtained on test–retest of the object substitution with conventional toys and imitated actions measures. There was no significant difference between test and retest scores for these measures based on a Wilcoxon Matched Pairs Signed-Ranks Test (Wilcoxon Sign Test).

CONCLUSION. Elaborate play scores, object substitution with conventional toys score, and imitation scores on the ChIPPA showed stability over time. Object substitution scores using unstructured materials were the least stable play measures and appeared to be related to the child's play themes. Since play is the primary occupation of children, it is essential that therapists have a reliable measure of play behavior. The test–retest reliability results from the ChIPPA provide evidence that this assessment produces a stable measure of play behavior that can then guide therapists when planning intervention strategies for children.


Play has been regarded as a phenomenon that changes from situation to situation (Barnett, 1991; Bledsoe & Shepherd, 1982; Sturgess & Ziviani, 1995). For example, a child’s play can vary according to whether the child is at home, at school, in a familiar setting, or the arrangement of a room (Barnett; McCune Nicolich, 1977). The perceived variability of play has contributed to the difficulties of knowing what to assess as play and has contributed to “existing play assessments lacking in adequate reliability and validity” (Sturgess & Ziviani, p. 108). In more recent times, reliable and valid play assessments that measure a child’s play in a child’s familiar environment have been developed (Bundy, 2001; Knox, 1997). However, there has been a lack of play assessments designed for use in occupational therapy clinical settings. At present, play assessments for use in a child’s familiar environment are rarely used in clinical practice or as part of assessment batteries (Couch, Dietz, & Kanny, 1997; Rodger, 1994). The theoretical importance of play as the primary occupation of childhood is therefore not reflected in the use of play assessment in assessment batteries. Stagnitti (2002a) has developed a clinical assessment of play called the ChIPPA. Stagnitti et al. (2000) have reported a preliminary study on the reliability and validity of this assessment.
The stability of the results from a test over time, known as test–retest reliability, is an important feature of a new assessment (Anastasi & Urbina, 1997; Kaplan, 1996). DeVellis (1991) prefers to use the term “temporal stability” rather than “test–retest reliability,” as the former indicates the extent to which the assessment and the underlying construct are stable over time. In test–retest studies both the measurement and the underlying construct (latent variable) are examined (DeVellis). Therefore if a poor result is obtained in test–retest studies, interpretation of the results involves deciphering whether the phenomenon itself has been subject to change or whether the measure is inherently unreliable (DeVellis).

The aim of this study was to establish the test–retest reliability of the ChIPPA (Stagnitti, 2002a; Stagnitti et al., 2000). Establishing test–retest reliability of this assessment allowed for examination of the changes in the underlying construct of self-initiated pretend play over time as well as documenting the reliability of the assessment.

Method

Participants

The sample comprised 38 preschool children ages 4 and 5 years old. Within the sample there were 23 boys and 15 girls. Nine boys and 9 girls were aged 4 years, 14 boys and 6 girls were aged 5 years. The children were recruited from two nursery schools and two preschool day care centers in a regional area. Of the 38 children in the study, 4 had reported developmental difficulties as assessed by a short developmental checklist based on the Revised Gesell Developmental Schedules (Knobloch, Stevens, & Malone, 1980) and 34 were typically developing children.

Instrument

The purpose of the ChIPPA (Stagnitti, 2002a) is to assess a child’s ability to self-initiate and organize pretend play. The ChIPPA is a standardized 30-minute assessment comprising one 15-minute session during which the child is assessed while playing with conventional toys and one 15-minute session during which the child is assessed while playing with unstructured play materials. The use of both conventional toys and unstructured play materials reflects the type of play that is being measured, as pretend play is defined as encompassing both conventional imaginative play and symbolic play. Conventional imaginative play occurs when the child uses conventional toys (that is, commercial toys) to pretend. For example, a child can pretend a doll is sick. Symbolic play involves the use of symbols in play. Lewis, Boucher, and Astell (1992) defined symbolic play as play involving substitution of one object to represent another (e.g., a block is used as a telephone); the attribution of a property to an action or object (e.g., the block is ringing); and reference to an absent object (e.g., the wave of an arm represents a window). The definition of pretend play assumes that the child can self-initiate these attributes of play.

Development of the ChIPPA was based on the assumptions underlying cognitive developmental theories of play. Such theories of play describe play as a voluntary activity that is a cognitive process, and propose that divergent thinking, adaptability to change, and flexibility are developed through play (Parham & Primeau, 1997; Piaget, 1962; Sutton-Smith, 1967). Within the cognitive developmental theories, play is regarded as an important behavior. This view is consistent with the emerging view in occupational therapy that play is important in itself and is a child’s primary occupation (Parham & Primeau; Rodger & Ziviani, 1999). The perspective of proponents of cognitive developmental theories is that pretend play contributes to a child’s cognitive and social development in a unique fashion compared to other forms of play behavior. Pretend play increases in complexity as a child develops, therefore, it is assumed that pretend play assessment can discriminate the play of developmentally younger children from developmentally older children.

Prior to administration of the ChIPPA, the play area is prepared by throwing a sheet over two adult-sized chairs placed side by side to make a “cubby” house. The ChIPPA is administered with the examiner and child sitting on the floor in front of the cubby house. Throughout the administration of the ChIPPA the examiner uses a nondirective approach similar to that expounded by Axline (1947). Table 1 summarizes the play materials and the administration of the ChIPPA.

Three play items are measured on the ChIPPA. These are the elaborateness of child’s play action (called Percentage of Elaborate Pretend Play Actions [PEPA]), the ability of the child to substitute objects during play (called Number of Object Substitutions [NOS]), and whether the child imitated the examiner’s modelled actions (called Number of Imitated Actions [NIA]). Since the ChIPPA has two sets of play materials (conventional toys and unstructured materials) and each is examined separately and as a combined (total) score, there are nine play measures on the ChIPPA. These measures are explained in Table 2.

An elaborate play action is defined as “a pretend play action that is executed carefully in a play context, theme, or sequence with attention to detail where the child uses controlled, fluid movements or talks through the play situation with attention to detail” (Stagnitti, 2002a, p. 184). (A different version of the ChIPPA for children with physical dis-
### Table 1. Summary of the Administration and Play Materials of the ChIPPA When Conventional Imaginative Play Session is Presented First

<table>
<thead>
<tr>
<th>ChIPPA Session</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional imaginative play session</strong>&lt;br&gt;(toys: truck, trailer, small boy doll, small girl doll, animals, fences, wrench)</td>
<td></td>
</tr>
<tr>
<td><strong>First 5 minutes</strong></td>
<td>Present the toys, except for the girl doll that is placed behind the examiner's back. No play directions are given. Suggested instruction: &quot;Here are some toys to play with. You can do what you want with them.&quot;</td>
</tr>
<tr>
<td><strong>Second 5 minutes</strong></td>
<td>Examiner produces the girl doll from behind his or her back and models five play actions. These actions are: walk doll, wave doll's hand, doll pats cow, doll drives truck, doll fixes wheels or fences with the wrench.</td>
</tr>
<tr>
<td><strong>Third 5 minutes</strong></td>
<td>The examiner stops modelling. No instructions are given. The child is encouraged to play by comments such as &quot;Can you think of anything else to play?&quot; When the time is up the examiner says, &quot;Time to stop playing with these toys. Now I've got some other toys to play with. Let's pack up so we can play with the new toys.&quot;</td>
</tr>
<tr>
<td><strong>Symbolic play session</strong>&lt;br&gt;(play materials: 1 shoe box, 1 larger box, tin, dowel stick, flat stick, 3 pebbles, 1 plastic cone, 1 tea towel, 1 face washer, 2 cloth &quot;dolls&quot;)</td>
<td></td>
</tr>
<tr>
<td><strong>First 5 minutes</strong></td>
<td>Present the unstructured play materials, except for one cloth &quot;doll.&quot; This cloth &quot;doll&quot; is left behind the examiner's back. No play directions are given. Suggested instruction: &quot;These toys are really different. You can do lots of things with these. You can play whatever you want with these toys.&quot;</td>
</tr>
<tr>
<td><strong>Second 5 minutes</strong></td>
<td>Examiner produces the cloth &quot;doll&quot; from behind his or her back and models five play actions. These actions are: wave &quot;doll's hand,&quot; place &quot;doll&quot; in box and place cloth over &quot;doll&quot; (as in going to bed), give &quot;doll&quot; a drink from the cone or the tin, place &quot;doll&quot; in box and drive the &quot;doll,&quot; walk &quot;doll.&quot;</td>
</tr>
<tr>
<td><strong>Third 5 minutes</strong></td>
<td>The examiner stops modelling. No instructions are given. The child is encouraged to play by comments such as &quot;Can you think of anything else to do with the objects?&quot; When the play session is over, the examiner says: &quot;That's all the time we have to play. Time to pack up now.&quot;</td>
</tr>
</tbody>
</table>

### Table 2. Summary of the Child-initiated Pretend Play Assessment Play Measures

<table>
<thead>
<tr>
<th>Play Score</th>
<th>Play Score Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elaborate Play Scores (PEPA)</strong></td>
<td></td>
</tr>
<tr>
<td>PEPA score for the conventional imaginative play session.</td>
<td>PEPA conventional</td>
</tr>
<tr>
<td>PEPA score for the symbolic play session.</td>
<td>PEPA symbolic</td>
</tr>
<tr>
<td>PEPA score for the conventional imaginative play session plus the symbolic play session.</td>
<td>PEPA combined</td>
</tr>
<tr>
<td><strong>Object Substitution Scores (NOS)</strong></td>
<td></td>
</tr>
<tr>
<td>NOS score for the conventional imaginative play session.</td>
<td>NOS conventional</td>
</tr>
<tr>
<td>NOS score for the symbolic play session.</td>
<td>NOS symbolic</td>
</tr>
<tr>
<td>NOS score for the conventional imaginative play session plus the symbolic play session.</td>
<td>NOS combined</td>
</tr>
<tr>
<td><strong>Number of Imitated Actions (NIA)</strong></td>
<td></td>
</tr>
<tr>
<td>NIA score for the conventional imaginative play session.</td>
<td>NIA conventional</td>
</tr>
<tr>
<td>NIA score for the symbolic play session.</td>
<td>NIA symbolic</td>
</tr>
<tr>
<td>NIA score for the conventional imaginative play session plus the symbolic play session.</td>
<td>NIA combined</td>
</tr>
</tbody>
</table>

**PEPA**—Percentage of Elaborate Pretend Play Actions  
**NOS**—Number of Object Substitutions  
**NIA**—Number of Imitated Actions
abilities is being studied in 2003.) The PEPA score reflects both the complexity and level of self-organization of a child's play ability. The scoring of PEPA is based on a child's actions, not intention, during play. This is an important distinction because the intention of the child cannot be known, whereas the action of the child can be observed and recorded. In the ChIPPAs, both play and nonplay actions are defined according to a Functional Verb List and Behavioural Verb List (Stagnitti, 2002b). These lists were based on a study by Jeffree and McConkey (1976). Knowing what to score as play is a crucial feature of a reliable assessment.

PEPA is calculated as the number of elaborate play actions divided by the total number of actions (play and nonplay) performed by the child during each 15-minute session of the ChIPPAs, multiplied by 100. Put in a formula, the PEPA score for the conventional imaginative play session is:

\[
\text{Number of pretend play actions} \times 100
\]

\[
\frac{\text{(conventional imaginative play session)}}{\text{Total number of actions during the conventional imaginative play session}}
\]

The NIA is scored in the middle 5-minute segment of each 15-minute session of the ChIPPAs when modelling by the examiner occurs. A score is recorded each time a child imitates one of the five modelled acts of the examiner.

NOS is calculated by counting the number of objects used in each object substitution. For example, if a child made a train with five objects, the score for NOS would be 5. The parameters to distinguish between manipulation of objects and object substitution have been defined in Stagnitti (2002b).

The ChIPPAs are scored from videotape (κ = .96 to κ = 1.0). Interrater reliability studies have still to be carried out in a clinical situation. In a validity study, the ChIPPAs performance, PEPA combined and NOS combined, were found to significantly discriminate (p < .001) between a sample of preschool children with preacademic problems and typically developing preschool children (Stagnitti et al., 2000). A significant difference (p < .001) was also found between a sample of typically developing children and developmentally delayed children on the combined measures of the ChIPPAs (Stagnitti, 2001). The use of the ChIPPAs in occupational therapy practice has led to the development of a treatment program for children called the Learn to Play program (Stagnitti, 1998).

**Procedure**

Ethical approval and consent for the research were obtained from the nursery schools, preschool day care centers, parents, and the university that sponsored the research. The ChIPPAs was administered to each child twice, 2 weeks apart, by the first author. The children were scored on the ChIPPAs as it was being administered. The children were seen individually in a separate room at either their nursery school or preschool day-care center. As one of the preschools in the study did not have a separate room, parents brought their children to a room at a local church. Therefore, for each child, the ChIPPAs test and retest were carried out in the same environment and under the same conditions.

As the same rater assessed the children each time, experimenter bias was a possibility (Portney & Watkins, 1993). To decrease bias, the rater did not reacquaint herself with the children's scores from the first assessment when conducting the retest. Bias was also reduced by the objective nature of the ChIPPAs scoring procedure. The ChIPPAs subsections were administered in the same order for the two assessments for each child, that is, a child would be given the conventional toys to play with first in both the test and retest. Instructions for administration, materials, and scoring of the ChIPPAs were strictly adhered to each time (Stagnitti, 2002b).

**Data Analysis**

An intraclass correlation coefficient (ICC) (Type 2,1) (Shrout & Fleiss, 1979) was used to measure test–retest reliability since this coefficient reflects degree of correspondence and agreement between scores on the first assessment and the retest (Portney & Watkins, 1993). As there are six different equations for calculating ICCs, the ICC chosen for this study used single ratings and was based on Model 2 of Shrout and Fleiss's three-model explanation (Portney & Watkins; Shrout & Fleiss). Model 2 was chosen because “rater” was considered a variable in this study as the consistency of the rater in the administration and scoring of the ChIPPAs over the 2 weeks was considered a variable as well as were differences between subjects and error variance over time. Further support for Model 2 was that the rater in this study did not have a separate room, parents brought their children to a room at a local church. Therefore, for each child, the ChIPPAs test and retest were carried out in the same environment and under the same conditions.
study was considered representative of a larger population of similar raters of the ChIPPA (Portney & Watkins). As the ICC is an inferential statistic, a normal distribution of scores is required. For data that violated normality, a Wilcoxon Matched Pairs Signed-Ranks Test (Daniel, 1999) and a percentage agreement were used. The Wilcoxon Matched Pairs Signed-Ranks Test was used to determine whether there was any significant difference between the test and retest scores. Percentage agreement was used to calculate the agreement between the scores. Results

Data for the three PEPA play measures and two of the NOS play measures (NOS symbolic and NOS combined) met the conditions for normality. Although NOS symbolic and NOS combined from the first administration of the ChIPPA had a Lilliefors significance less than .05 (.0437 and .0410, respectively), the kurtosis and skewness of the NOS symbolic and NOS combined distribution were within acceptable limits (Coakes & Steed, 1997). Hence, ICC statistics were calculated for these five play measures and they are reported in Table 3.

The remaining play measures (NOS conventional, NIA conventional, NIA symbolic, and NIA combined) had a restricted range in scores with the majority of children scoring 0 on both test and retest. For example, for NIA conventional, 32 of the 38 children scored 0 on each occasion, and for NIA symbolic 29 children out of the 38 children scored 0 on each occasion. The lack of variability within these particular scores on the ChIPPA affected the correlational analyses. Anastasi and Urbina (1997) noted that if the range of the individual differences in a group is small, then the correlation will be close to 0. This was the case with the play measures NOS conventional, NIA conventional, NIA symbolic, and NIA combined. The use of a weighted Kappa was not suitable for these data. In this study there was one rater and low variability between scores on the test and retest (Portney & Watkins, 1993). Typical 4–5-year-old children do not rely on the modelled actions of the examiner to play and tend to play with the conventional toys conventionally. Thus, it is argued that the high incidence of 0 for the test and retest for imitated actions and conventional toy object substitution scores is due less to error and more to the typical development of preschoolers. Consequently, a Wilcoxon Matched Pairs Signed-Ranks Test was carried out to ascertain whether there was any significant difference between the test and retest groups. No groups were significantly different. To ascertain the agreement between the scores from the first assessment and the retest for these play items, a percentage agreement was calculated. Results for the percentage agreement ranged from 63.2% to 84.2%.

Discussion

The Temporal Stability of the ChIPPA

The results of the ICCs, Wilcoxon Matched Pairs Signed-Ranks Test, and the percentage of agreement indicated that the ChIPPA possesses temporal stability. Using Portney and Watkins’ (1993) general guideline for determining test–retest reliability, PEPA symbolic and PEPA combined had ICCs above .75 indicating good reliability. PEPA conventional, NOS symbolic, and NOS combined had ICCs between .56 and .73 indicating moderate reliability. PEPA conventional was the most stable play measure and NOS symbolic was the least stable measure.

For the remaining variables, NOS conventional, NIA conventional, NIA symbolic, and NIA combined, results indicated that there were percentage agreements above 63% and no significant difference between test and retest scores of these play measures. These measures have a narrow distribution of scores. The lack of variability in NOS conventional is due to the tendency of children to use the conventional play materials in a conventional manner rather than a symbolic manner. However, this play measure is being retained in the ChIPPA because a child’s use of the conventional toys in object substitution is indicative of an advanced object substitution developmental skill level. The narrow distribution of the NIA play measures is due to the

### Table 3. Intraclass Correlation Coefficients (Type 2,1) for the Three PEPA Play Measures and NOS Symbolic and NOS Combined on the ChIPPA

<table>
<thead>
<tr>
<th>Play Measure</th>
<th>Test</th>
<th>Retest</th>
<th>ICC(2,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEPA conventional</td>
<td>M 64.18</td>
<td>M 64.63</td>
<td>.729*</td>
</tr>
<tr>
<td>SD 18.55</td>
<td>SD 17.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEPA symbolic</td>
<td>M 53.54</td>
<td>M 51.17</td>
<td>.771**</td>
</tr>
<tr>
<td>SD 26.73</td>
<td>SD 28.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEPA combined</td>
<td>M 117.72</td>
<td>M 115.80</td>
<td>.845**</td>
</tr>
<tr>
<td>SD 42.14</td>
<td>SD 41.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOS symbolic</td>
<td>M 11.89</td>
<td>M 11.00</td>
<td>.556*</td>
</tr>
<tr>
<td>SD 9.48</td>
<td>SD 9.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOS combined</td>
<td>M 11.97</td>
<td>M 11.16</td>
<td>.572*</td>
</tr>
<tr>
<td>SD 9.58</td>
<td>SD 10.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* moderate correlation (Portney & Watkins, 1993)
** good correlation (Portney & Watkins, 1993)

N = 38

Key:
PEPA—Percentage of Elaborate Pretend Play Actions
NOS—Number of Object Substitutions
Conventional—Conventional imaginative session
Symbolic—Symbolic play session
Combined—Conventional imaginative session plus symbolic play session
fact that typically developing preschool children, who comprised the majority of children in the sample, do not tend to imitate the modelled play actions because they are capable of self-initiating their own play ideas. Stagnitti et al. (2000) noted that some children with preacademic problems who had low scores in elaborate play actions and object substitutions, also had low scores in imitated actions. They suggested that this may have been due to the fact that the play of these children was so deficient that the children did not recognize the modelled actions of the examiner as cues on how to use the play materials. Children with developmental delay have been found to exhibit high scores on NIA measures because they relied on a model for play ideas (Stagnitti, 2001).

Object substitution during the symbolic play session and the NOS combined score were the least stable measures. Change in the underlying construct of the substitution of objects is put forward as the reason why the play measures NOS symbolic and NOS combined had lower reliability than the elaborate play, object substitution with conventional toys, and imitated action measures. The lower test–retest reliability of these NOS measures may have been due to the change in play theme that the children used during the two assessments. This shift had the effect of changing the uses of the objects for the play theme and affected the number of object substitutions used during play. NIA combined had the lowest percentage agreement of the remaining four measures. The symbolic play materials are developmentally more challenging to play with compared to the conventional toys. Some children imitated actions during the first test and other children imitated during the retest. Proposed reasons for this are that children who imitated during the first test used the modelled actions for play ideas whereas children who imitated on the retest were trying these play actions for the first time.

Given that the test and retest results showed that the ChIPPA was a stable measure of play behavior, it is suggested that the underlying construct (or latent variable) of the ChIPPA child-initiated pretend play influenced the observed scores of the ChIPPA more than the extraneous influences from the environment (DeVellis, 1991). Other studies of play measurement have found that self-initiated play can vary in regard to extraneous influences in the environment (Bledsoe & Shepherd, 1982; Sturgess & Ziviani, 1995). Examples of extraneous influences noted during this study were children remembering the play materials from the first administration of the ChIPPA and children discussing the play assessment with each other. For example, some children commented “I remember these. Let’s do what I did last time. Remember,” or “I’ll do what I did last week.” Discussion among children was evidenced by comments such as “(child’s name) said he built a slide. How did he do that?” One child requested new toys on the retest as he had already played with the toys 2 weeks before. Extraneous influences such as changes in environment since last assessment were also evident. For example, between the test and retest there had been a magic show on television. One child explained his magic tricks in a similar fashion to the television version.

The results lend support to the stability of the ChIPPA as a measure of play. The use of standardized play materials, instructions, and scoring contributed to the stability of the ChIPPA. Assessing the children in the same physical environment for the test and retest of the ChIPPA also contributed to the results.

Study Limitations
When deciding on the time frame between administrations of the ChIPPA assessment, consideration was given to the construct assessed as well as the subjects’ memories, progressive developmental changes, and the opportunity to discuss play ideas with others in the intervening time period. This study used a 2-week time interval so the influence of progressive developmental changes was reduced and, for a preschooler, a 2-week time frame was thought to be long enough for memory of the play situation to fade. A 2-week period is common in studies with young children (Anastasi & Urbina, 1997). However, several children did remember what they had played the time previously or had discussed the play with the other children.

Conclusion
The ChIPPA was found to meet good reliability levels in the elaborate measures PEPA symbolic and PEPA combined. Elaborate play measures are the most complex measures to score on the ChIPPA. The three elaborate play measures were stable with ICCs above .72. The play measures NOS symbolic and NOS combined were the least stable measures and it was argued that object substitution changed as a function of the child’s play theme. For the remaining measures, a percentage agreement and Wilcoxon Matched Pairs indicated no significant difference between the groups and percentage agreements were above 63%. It is recommended that further studies of the test–retest reliability of the ChIPPA be carried out to explore further the measure of object substitution and stability of the ChIPPA measures. The standardization of the play materials, administration, and scoring of the ChIPPA contributed to the test–retest reliability results.
Acknowledgments

This study was supported by a LaTrobe University Postgraduate Research Scholarship. A paper on the test–retest reliability of the Child-Initiated Pretend Play Assessment was presented at the OT Australia 21st National Conference held in Brisbane, 2001. This study is based on work presented in the PhD thesis of the first author. We wish to thank Dr. Peter Gibbon for his critical appraisal during preparation of the manuscript.

References