Objectivity and Stability of the Preschool Imitation and Praxis Scale

Marleen Vanvuchelen, Herbert Roeyers, Willy De Weerdt

KEY WORDS
- child, preschool
- imitative behavior
- psychomotor performance
- reproducibility of results

OBJECTIVE. We examined rater and test–retest reliability of the Preschool Imitation and Praxis Scale (PIPS).

METHOD. We administered the PIPS to 119 typically developing children ages 1.5–4.9 yr.

RESULTS. The PIPS demonstrated acceptable intra- and interrater reliability on item level ($k_w = 0.45–1$) and scale level (intraclass correlation coefficient ICC = 0.996; 95% CI: 0.968–0.999 and ICC = 0.995; 95% CI: 0.990–0.997, respectively). The smallest detectable difference of the PIPS was 5.6%, indicating that the change score rated by different raters for an individual child is valid and that the PIPS can be used by different raters as an outcome measure to determine children's improvement or maturation. Results of test–retest analysis revealed that the PIPS score is stable over time ($r = .93$).

CONCLUSION. The PIPS appears to meet the required standards regarding objectivity and stability. The PIPS may assist clinicians and researchers in evaluating and reevaluating preschoolers’ imitation ability, which is a primary learning strategy of young children.


Imitation is often thought of as a low-level, cognitively undemanding, even childish form of behavior. However, recent work across a variety of sciences has argued that imitation is an extraordinary ability that is fundamentally linked to characteristically human forms of intelligence. We use imitation for learning motor skills, for facilitating comprehension of other people’s actions and mental states, and as a communicative reference to actions or people (Hurley & Chater, 2005).

The term imitation has many definitions but is most commonly used in two connotations. The first connotation defines imitation as the capacity of an individual to replicate an observed motor act. The observer’s perception of the demonstrator’s behavior causes similar behavior in the observer (Prinz, 2002). The second connotation defines imitation as the capacity to acquire, by observation, a new motor behavior and to repeat it using the same movements used by the demonstrator (Tomasello, Carpenter, Call, Behne, & Moll, 2005). In both cases, imitation requires the capacity to transform perceptual information into a motor copy of it.

In child development research, imitation is seen as a window into the sensorimotor, cognitive, and social abilities of young children, including those with atypical development (Rogers & Williams, 2006). Imitation problems are frequently reported in children with autism spectrum disorders (Rogers & Williams, 2006; Vanvuchelen, 2009; Vanvuchelen, Roeyers, & De Weerdt, 2007a, 2007b, 2011b, 2011c; Williams, Whiton, & Singh, 2004), with intellectual disabilities (Macedoni-Luksic et al., 2009) and, to a lesser extent, in children with developmental coordination disorders (Green et al., 2002). Typically developing young children are remarkably adept at imitating and do
so with the apparent motivation not just to solve a task but also to join the social interaction (Tomasello et al., 2005).

In young children, imitation is essential to learning social behavior and skilled acts or praxis (Masur, 2006). Imitation avoids time-consuming trial-and-error learning. In reproducing the exact and detailed features of the demonstrator’s actions, children are likely to successfully complete the intended actions, even with a limited understanding of their purpose. Moreover, faithful copying can be used to disentangle the goal of an action to be imitated when it is not completely clear to the child or to learn about initially opaque aspects of causality (Lyons, Young, & Keil, 2007). Consequently, the accuracy of the perception–action coupling to match the visual kinematic features of a perceived action to the motor kinematic features of the child’s own action is an important criterion for children’s aptitude to learn new skills.

Because imitation is essential for young children to learn skilled acts, a test to identify young children with delayed imitation is important. Imitation is frequently implicitly assessed in the context of standardized cognitive and motor developmental testing. For instance, the Mullen Scales of Early Learning (Mullen, 1995), the Bayley Scales of Infant and Toddler Development, 3rd Edition (Bayley, 2006), and the Peabody Developmental Motor Scales, 2nd Edition (Folio & Fewell, 2000) have many items that require demonstration by an examiner in addition to, or in lieu of, verbal instructions. In clinical settings, pediatric occupational therapists, psychologists, and physical therapists frequently use four published measures to assess imitation abilities in young children: the Sensory Integration and Praxis Test (Ayres, 1989), the Test d’Imitation de Gestes (Bergès & Lézine, 1963), the Developmental Neuropsychological Assessment (Korkman, Kirk, & Kemp, 1997), and the Uzgiris–Hunt Scales of Psychological Development (Uzgiris & Hunt, 1987). Although these measures (reviewed in Table 1) rely on a theoretical framework and prove consistency of the scores, they suffer from at least one of the following limitations: (1) They lack age-specific tasks for preschoolers, (2) they consider imitation to be a one-dimensional construct and are limited to bodily imitation tasks, (3) their construct validity regarding distinct domains of imitation was never investigated, and (4) they do not provide comparative normative data of typically developing preschoolers.

To address these issues, we developed the Preschool Imitation and Praxis Scale (PIPS; Vanvuchelen, 2009), which was designed to assess the accuracy of bodily (i.e., actions without objects) and procedural (i.e., actions with objects) imitation performances in young children. The theoretical framework and deductive test construction approach of the PIPS have been described in detail elsewhere (Vanvuchelen et al., 2011a). To recap briefly, different action types considered to be important as revealed unraveled in research (Petreska, Adriani, Blanke, & Billard, 2007) in apraxic adults were selected to tap a broad range of possible imitation mechanisms: action types with different effects (salient environmental and internal), representational levels (meaningful and nonmeaningful; goal directed and not goal directed), temporal complexities

<table>
<thead>
<tr>
<th>Supporting Reference</th>
<th>Measure</th>
<th>Age Range</th>
<th>Theoretical Framework</th>
<th>Reliability</th>
<th>Age-Specific Tasks for Preschoolers</th>
<th>Construct Validity</th>
<th>Action Types</th>
<th>Tasks (No. of Tasks)</th>
<th>Standardization (Sample Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayres (1989)</td>
<td>Sensory Integration and Praxis Test</td>
<td>4 yr–8 yr, 11 mo</td>
<td>Sensory Integration Theory</td>
<td>Interrater Test–retest</td>
<td>No</td>
<td>One domain</td>
<td>Bodily</td>
<td>si-NM-postural (17) sq-NM-gestural (9) Oral (19)</td>
<td>2,000</td>
</tr>
<tr>
<td>Bergès &amp; Lézine (1963)</td>
<td>Test d’Imitation de Gestes</td>
<td>3 yr–6 yr, 11 mo</td>
<td>Neuropsychological theory</td>
<td>Interrater Test–retest</td>
<td>Yes</td>
<td>One domain</td>
<td>Bodily</td>
<td>si- and bi-NM-gestural (36)</td>
<td>489</td>
</tr>
<tr>
<td>Korkman, Kirk, &amp; Kemp (1998)</td>
<td>Developmental Neuropsychological Assessment</td>
<td>3 yr–4 yr, 11 mo</td>
<td>Assessment of adults with acquired brain damage</td>
<td>Interrater Test–retest Split-half</td>
<td>Yes</td>
<td>One domain</td>
<td>Bodily</td>
<td>si-NM-gestural (5)</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Note: NM = nonmeaningful, si = single; sq = sequential; bi = bimanual.
birthweight over time. The test–retest reliability, or reproducibility, provides an indication of the stability of the child's imitation ability (van Baalen, Odding, van Woensel, & Roebroeck, 2006). Kropmans, Dijkstra, Stegenga, Stewart, & de Bont, 1999; van Baalen, Odding, van Woensel, & Roebroeck, 2006). The test–retest reliability, or reproducibility, provides an indication of the stability of the child’s imitation ability over time.

Method

Participants

To achieve a sample representative of a typical population of preschool children, we used a stratified random sampling procedure of daycare centers and regular preschools in Flanders, the Dutch-speaking region within Belgium. The sample consisted of 119 typically developing preschool children (69 girls and 50 boys) ages 1.5–4.9 yr.

Criteria for admission into the study were that the children were not born preterm (>36 wk gestation age and birthweight >1,500 g) and had no known physical, sensory, or mental handicap according to the parental report in the Ages and Stages Questionnaires (ASQ; Bricker & Squires, 1999). The ASQ is a parent-administered structured questionnaire that covers five domains of child development: (1) communication, (2) gross motor, (3) fine motor, (4) problem solving, and (5) personal–social skills. The ability of the ASQ system to correctly identify typically developing children (specificity) is high: 86%. The ability to detect delayed development (sensitivity) is, on average, 72% (Bricker & Squires, 1999). The mean time interval between the parental report and the test was 3 wk.

The socioeconomic status (SES) of the children was determined by the educational level of mother and father expressed in educational years: Level 1 (<7 yr), Level 2 (7–10 yr), Level 3 (11–12 yr), Level 4 (13–16 yr), and Level 5 (>16 yr). The combined educational level of both parents, expressed as the sum score of both SES scores, was used as an indicator for the child’s SES. The distribution of the combined SES scores was as follows: Score 3 = 1%; Score 4 = 1%; Score 6 = 16.7%; Score 7 = 12.5%; Score 8 = 13.5%; Score 9 = 16.7%, and Score 10 = 38.6%.

The parents of all children provided informed consent before their child’s participation.

Instrument

The PIPS consists of 10 task categories: 6 gestural, 3 procedural, and 1 facial. The 10 task categories and 30 PIPS tasks are described in Table 2. The six task categories of gestural imitation are as follows:

1. Meaningful intransitive gestures (i-MG; e.g., communicative gestures such as “perform the gesture to wave good-bye”)
2. Meaningful goal-directed transitive gestures (t-MG; e.g., “pretend to comb your hair with an imaginary comb”)
3. Nonmeaningful single-hand postures (si-NMG; e.g., “raise your outstretched arm till 90° anteflexion and make a fist”)
4. Nonmeaningful bimanual hand postures (bi-NMG; e.g., “place one fist on top of the other”)
5. Nonmeaningful hand postures to the face and head (fa-NMG; e.g., “touch the top of your nose with the extended index finger”)
6. Nonmeaningful sequences of hand postures (sq-NMG; e.g., “hit the table with the palm of your hands, cross the arms, and hit the table again, return to the original position, and hit the table once more”).

The three task categories of procedural imitation are as follows:

1. Goal-directed substituted actions on objects (sao-P; e.g., “raise a toy bear by pulling a cord”)
2. Goal-directed actions on substituted objects (aso-P; e.g., “turn a cup upside-down and play drums on it with two spoons”)
3. Non–goal-directed action sequences on objects (sq-P; e.g., “open a box, put the lid on the table, turn the box upside-down, put a block on the bottom of the box”).

There is one facial imitation category (f; e.g., “shake the head, eyes closed to say ‘no’ with an expression of disapproval”). Facial expressions have a communicative meaning.
Table 2. Description of the 30 Items of the Preschool Imitation and Praxis Scale (PIPS) Presented in Standardized Order

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Type</th>
<th>Category</th>
<th>Dimension</th>
<th>Task Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P</td>
<td>sao-P1</td>
<td>GDP</td>
<td>Raise a toy bear by pulling a cord.</td>
<td>0-2</td>
</tr>
<tr>
<td>2</td>
<td>P</td>
<td>sao-P2</td>
<td>GDP</td>
<td>Put a wooden block on top of your head.</td>
<td>0-2</td>
</tr>
<tr>
<td>3</td>
<td>P</td>
<td>sao-P3</td>
<td>GDP</td>
<td>Switch on a lamp in a toy animal with your forehead.</td>
<td>0-2</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>i-MG1</td>
<td>SIB</td>
<td>Perform the gesture to “wave good-bye.”</td>
<td>0-2</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>i-MG2</td>
<td>SIB</td>
<td>Perform the gesture to “show something with your finger.”</td>
<td>0-2</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>i-MG3</td>
<td>SIB</td>
<td>Perform the gesture to “beckon with the index finger.”</td>
<td>0-2</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>si-NMG1</td>
<td>SIB</td>
<td>Raise your outstretched arm till 90° anteflexion and make a circle with your fingers and thumb.</td>
<td>0-3</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>si-NMG2</td>
<td>SIB</td>
<td>Raise your outstretched arm till 90° anteflexion and stretch out your fingers.</td>
<td>0-3</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>si-NMG3</td>
<td>SIB</td>
<td>Raise your outstretched arm till 90° anteflexion, hold up the little finger while all the other fingers and the thumb are bent.</td>
<td>0-3</td>
</tr>
<tr>
<td>10</td>
<td>P</td>
<td>sq-P1</td>
<td>NGDP</td>
<td>Open the box, put the lid on the table, turn the box upside-down, put the block on the bottom of the box.</td>
<td>0-3</td>
</tr>
<tr>
<td>11</td>
<td>P</td>
<td>sq-P2</td>
<td>NGDP</td>
<td>Take the block from the bottom of the box, turn the box in normal position again, close the box, put the block on the lid of the box.</td>
<td>0-3</td>
</tr>
<tr>
<td>12</td>
<td>P</td>
<td>sq-P3</td>
<td>NGDP</td>
<td>Take the block from the lid of the box, open the box, put a disc into the box, close the box, put the block again on the lid of the box.</td>
<td>0-3</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>t-MG1</td>
<td>SIB</td>
<td>Pretend to “comb your hair with an imaginary comb.”</td>
<td>0-4</td>
</tr>
<tr>
<td>14</td>
<td>G</td>
<td>t-MG2</td>
<td>SIB</td>
<td>Pretend to “open an imaginary door with an imaginary key.”</td>
<td>0-4</td>
</tr>
<tr>
<td>15</td>
<td>G</td>
<td>t-MG3</td>
<td>SIB</td>
<td>Pretend to “brush your teeth with an imaginary toothbrush.”</td>
<td>0-4</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>bi-NMG1</td>
<td>SIB</td>
<td>Place one fist on top of the other.</td>
<td>0-3</td>
</tr>
<tr>
<td>17</td>
<td>G</td>
<td>bi-NMG2</td>
<td>SIB</td>
<td>Extend the index fingers of both hands while the other fingers and thumbs are bent, and bring the top of the index fingers toward each other.</td>
<td>0-3</td>
</tr>
<tr>
<td>18</td>
<td>G</td>
<td>bi-NMG3</td>
<td>SIB</td>
<td>Open one hand in vertical position and touch the top of the fingers with the palm of the other hand in horizontal position.</td>
<td>0-3</td>
</tr>
<tr>
<td>19</td>
<td>G</td>
<td>fa-NMG1</td>
<td>SIB</td>
<td>Extend your index finger and touch the top of your nose.</td>
<td>0-3</td>
</tr>
<tr>
<td>20</td>
<td>G</td>
<td>fa-NMG2</td>
<td>SIB</td>
<td>Touch your lower lips with the nails of your thumbs.</td>
<td>0-3</td>
</tr>
<tr>
<td>21</td>
<td>G</td>
<td>fa-NMG3</td>
<td>SIB</td>
<td>Extend the index finger of your left hand and touch your right cheek, and extend the index finger of your right hand and touch your left cheek.</td>
<td>0-3</td>
</tr>
<tr>
<td>22</td>
<td>P</td>
<td>aso-P1</td>
<td>GDP</td>
<td>Turn a cup upside-down and play drums on it with two spoons.</td>
<td>0-2</td>
</tr>
<tr>
<td>23</td>
<td>P</td>
<td>aso-P2</td>
<td>GDP</td>
<td>Remove the cap of a doll and put a shoe on the head of the doll.</td>
<td>0-2</td>
</tr>
<tr>
<td>24</td>
<td>P</td>
<td>aso-P3</td>
<td>GDP</td>
<td>Put a toy car in bed, turn it upside-down, and tuck it in with a blanket.</td>
<td>0-2</td>
</tr>
<tr>
<td>25</td>
<td>G</td>
<td>sq-NMG1</td>
<td>SOB</td>
<td>Hit the table with the palm of your hands, cross the arms and hit the table again, return to the original position and hit the table once more.</td>
<td>0-3</td>
</tr>
<tr>
<td>26</td>
<td>G</td>
<td>sq-NMG2</td>
<td>SOB</td>
<td>Hit the table with one hand in supination, turn the hand in pronation and hit the table again, clap the hands, hit the table with the palm of both hands.</td>
<td>0-3</td>
</tr>
<tr>
<td>27</td>
<td>G</td>
<td>sq-NMG3</td>
<td>SOB</td>
<td>Hit the table with both hands in supination, turn the hands in pronation, hit the table again, clap the hands, hit the table with the palm of both hands once more.</td>
<td>0-3</td>
</tr>
<tr>
<td>28</td>
<td>F</td>
<td>f1</td>
<td>SIB</td>
<td>Shake the head, eyes closed to say “no,” with an expression of disapproval.</td>
<td>0-2</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>f2</td>
<td>SIB</td>
<td>Look angry with a frown of the eyebrows.</td>
<td>0-2</td>
</tr>
<tr>
<td>30</td>
<td>F</td>
<td>f3</td>
<td>SIB</td>
<td>Nod quickly with your head and show an expression of happiness.</td>
<td>0-2</td>
</tr>
</tbody>
</table>

Note. P = procedural imitation; G = gestural imitation; F = facial imitation; sao-P = substituted actions on objects; i-MG = intransitive meaningful gestures; si-NMG = single nonmeaningful hand postures; sq-P = action sequences on objects; t-MG = transitive meaningful gestures; bi-NMG = bimanual nonmeaningful hand postures; fa-NMG = nonmeaningful hand postures to the face or head; aso-P = actions on substituted objects; sq-NMG = sequences of nonmeaningful hand postures; f = facial expressions; GDP = goal-directed procedural imitation; SIB = single bodily imitation; NGDP = non–goal-directed procedural imitation; SOB = sequential bodily imitation.

Exploratory factor analysis disclosed four dimensions in the scale: goal-directed versus non–goal-directed procedural imitation and single versus sequential bodily imitation. PIPS scale and subscale scores have high internal consistency (Vanvuchelen et al., 2011a).

Imitation performances are scored on a 3- or 5-point scale in accordance with the criteria of the scoring system of the PIPS (Vanvuchelen, 2009), which evaluated the spatiotemporal resemblance between the modeled and copied action. To illustrate this system, we use the scoring of the task “pretend to comb your hair with an imaginary comb.” Score 3 is given if the child has used a symbolic grip and has performed a repetitive action on both sides of the head. Score 2 is given if the child has used a body-part-as-an-object grip and has performed a repetitive action on one side of the head or a single action. Score 1 is given if the child has used a body-part-as-an-object grip and has performed a repetitive action on one side of the head or a single action. Score 0 is given if the child has performed another action or has refused to imitate. The final PIPS score is a reflection of the accuracy of the child’s imitation performance (Vanvuchelen, 2009).
Procedure

The children were assessed by trained examiners. The PIPS training for the examiners was spread over 2 wk. During the first session of 2 hr, the trainer (Marleen Vanvuchelen) explained the item instructions and the scoring system, and a training video was observed. During the second session of 3 hr, the examiners were given feedback on their test administration, and three other training videos were scored independently. An interrater agreement of the total score >85% with the trainer was achieved by all examiners.

Each child was individually assessed in a quiet room and was seated at a table in front of the examiner. Before the 30 tasks of the PIPS were administered, the child was given three exercises: (1) the imitation of “removing five beads one by one from a string and putting them in a cup”; (2) “clapping the hands,” and (3) “raising an open hand.” During these introductory tasks, a broad range of instructions to evoke imitation was given to the child: demonstrations, verbal commands, physical assistance.

The administration of the 30 tasks of the PIPS was in accordance with the guidelines for item instruction of the PIPS (Vanvuchelen, 2009). The 30 tasks were presented in a standardized way; for example, left- and right-handed actions were demonstrated alternately. The child was free to imitate with the left or right hand. Before demonstrating each task, the child’s attention was attracted by addressing the child by name. Only the verbal instruction “____________ (name), you do this, too” was given. The time needed to complete the PIPS ranged from 10 to 20 min.

Data Analysis

Relative Intrarater and Interrater Reliability. To determine relative intrarater reliability, the imitation performance of a random selection of 21 of the 119 children (13 girls and 8 boys) 1.8–4.7 yr (mean [M] age = 37.8 mo, standard deviation [SD] = 11.7 mo) was videotaped. One rater scored the imitation performances twice: immediately during the assessment (live rating) and 4 mo later (video rating).

To establish relative interrater reliability, the imitation performance of a random selection of 42 children (25 girls and 17 boys) 1.8–4.8 yr (M = 38.3 mo, SD = 11.9 mo) was videotaped. Two trained examiners scored the performances on the videotapes independently.

Relative intrarater and interrater reliability at the individual item level were examined using Cohen’s weighted κ values. The interpretation of κ values was done according to Fleiss (1981): <0.40 = agreement by coincidence; 0.40–0.60 = moderate; 0.61–0.75 = good, >0.75 = excellent agreement. Percentages of agreement were also determined for all items: ≥70% was considered as acceptable.

Relative intrarater and interrater reliability on scale level were examined with the intraclass correlation coefficient (ICC). For the computation of the ICCs for the intrarater agreement, the option “treating rater as a fixed effect” was applied. Because only one rater was involved, a two-way mixed-effects model with measures of absolute agreement was used. For the computation of the ICCs for the interrater agreement, the option “treating participants as well as raters as a random effect” was applied. This strategy resulted in a two-way random effects model with measures of absolute agreement. Bland and Altman’s (1986) limits of agreement on the scale score (mean of the differences between ratings ±2 SD) were used to assess the strength of agreement within and between raters.

Absolute Interrater Reliability. Absolute interrater reliability was examined by the smallest detectable difference (SDD) to determine whether the change scores rated by different raters of an individual child were valid (beyond random errors) at the 95% confidence level. First, the standard error of measurement (SEM) for the two separate ratings (SEM<sub>first</sub> and SEM<sub>second</sub>) was calculated on the basis of the SD and the ICC:

\[
SEM = SD \times \sqrt{1 - ICC}
\]

which was used to quantify the amount of random measurement errors. Assuming that measurement errors are distributed normally, the corresponding 95% confidence interval (CI), in which the true score is expected, is ±1.96 × SEM. No clear criteria for an acceptable SEM value are available. Therefore, the extent to which ratings can deviate from one another will be a question of judgment that should preferably be defined in advance (Bland & Altman, 1986). We expressed the SEM as a percentage of the possible score range of the PIPS (0–81). We postulated that a value <10% (van Baalen, Oddenig, van Woensel, & Roebroeck, 2006) and a difference between the percentages of two SEMs <1% would be satisfactory.

Then, the SDD was calculated using the following formula (Kropmans et al., 1999):

\[
SDD = 1.96 \sqrt{SEM_{first}^2 + SEM_{second}^2}
\]

If the difference between both SEMs was <1%, we assumed that they were equal and simplified the formula as follows:
A statistically significant change between two ratings must be larger than the smallest detectable difference (Kropmans et al., 1999).

Test–Retest Reliability. For the test–retest reliability, 56 typically developing children (31 girls, 25 boys) ages 1.5–4.9 yr (M = 36.5 mo, SD = 11.9) were measured twice with a time-interval of 1 wk. Given that children may become familiar with the test items when tested a second time, two PIPS equivalent forms (PIPS Form A and PIPS Form B) were used. Half of the children were first examined with the PIPS Form A and 1 wk later with PIPS Form B. The other children were examined in the opposite order. Test–retest reliability was calculated using a Pearson product–moment correlation coefficient. A coefficient of .90 was set as a minimal requirement (Field, 2005).

Results

Relative Intrarater–Interrater Reliability

Table 3 provides the weighted κ values and percentages of agreement for the intrarater and interrater agreement of the 30 individual item scores.

Among the 30 PIPS items, 20 revealed an excellent intrarater agreement for individual item scores, 4 had good agreement, and 2 had moderate agreement. The κ statistics of 4 items could not be computed because of the skewed distribution of the data. However, the percentage of agreement of the following items was obtained: t-MG3, 76%; t-MG4, 90%; sq-NMG1, 57%; and sq-NMG2, 61%. The intrarater reliability of the PIPS total score was high (ICC = .996; 95% CI: 0.968–0.999).

Among the 30 PIPS items, 17 revealed an excellent interrater agreement for individual item scores and 10 had good agreement. The κ values of 3 items could not be calculated, because of the skewed distribution of the data. However, the percentage of agreement of these three items was obtained: t-MG3, 76%; t-MG4, 90%; sq-NMG1, 57%; and sq-NMG2, 61%. The intrarater reliability of the PIPS total score was high (ICC = .995; 95% CI: 0.990–0.997).

Figure 1 plots limits of agreement calculated from intra- and interrater data of the PIPS scale scores. Plotting PIPS means of the two observations against the intrarater and interrater differences for each participant does not give any indication that measurement errors vary systematically over the range of possible scores.

Absolute Interrater Reliability

The SEM of the first rater was 1.66 (2.02%) and of the second rater, 1.74 (2.12%). The difference between both was 0.1%. Because the assumption of equal SEMs was met, the SDD was calculated using the SEM of the first rater. The SDD was 4.6 (5.6% of the possible score range of the PIPS), indicating that the change score rated by different raters for an individual child is valid.

Test–Retest Reliability

The association between the scores of the 56 children on the similar forms (PIPS Form A and PIPS Form B) assessed with a time interval of 1 wk was high (r = .93, p < .001).
Discussion

Imitation is the ability to learn a novel task by replicating both the result and the manner of achieving it from demonstration. It plays a central role in learning without (yet) understanding the actions of the demonstrator. When young children observe an adult manipulating a novel object, they automatically encode all of the adult’s purposeful actions as causally necessary. They implicitly treat the adult’s actions as highly reliable indicators of the object’s causal structure (Lyons et al., 2007). Pediatric occupational therapists, physical therapists, and psychologists who are concerned with the education and intervention of young children are convinced of the importance of this learning strategy. Imitation problems are mainly reported in children with autism spectrum disorders (Rogers & Williams, 2006; Vanvuchelen et al., 2007a, 2007b, 2011b, 2011c; Williams et al., 2004) and intellectual disabilities (Macedoni-Luksic et al., 2009). An assessment to identify children with delayed imitation abilities is evidently important.

It is also essential to ensure that measurements developed for clinical and research purposes are adequately constructed, reliable, and valid. The main goal of this study was to examine the reliability of the PIPS, which was designed to assess the accuracy of bodily and procedural imitation performances in young children (Vanvuchelen, 2009). Discrepancy among repeated measurements can be caused by random and systematic errors (Rousson, Gasser, & Seifert, 2002).

Random errors reduce the objectivity of the data and can arise from inconsistencies in the administering and scoring protocol. Therefore, the examiners were trained in these protocols to prevent inconsistencies of scores. The intrarater and interrater reliability were, for the most part, high at the individual item level and excellent for the total score of the PIPS. Transitive meaningful gestures and sequences of nonmeaningful gestures were the two task categories that were most difficult to score in a reliable way. Therefore, the description of the scoring criteria of these tasks in the PIPS guidelines has been revised. The results of intrarater agreement revealed that the PIPS can be reliably scored by trained examiners without the use of video recording.

In clinical settings, it is essential to know by how much the score of one rater is likely to differ from the score of another rater. If this value is not enough to cause problems in clinical interpretation, one can use the two raters interchangeably. Interrater reliability based on correlation coefficients provides an indication of relative reliability (Atkinson & Nevill, 1998). A high ICC, however, does not always mean that the two raters agree. Correlation measures the strength of a relation between two scores, not the agreement between them. When data of two ratings are plotted, perfect correlation exists if the points lie along any straight line. However, there is perfect agreement only if the points lie along the line of equality (Bland & Altman, 1986). Because correlations are highly influenced by the range of values in the sample, it is essential to calculate the absolute reliability, including the SDD. If the PIPS is used in intervention and longitudinal studies, knowledge of the SDD is important for clinicians and researchers. The SDD is important in determining whether the change scores rated by different raters for an individual child after therapy or maturation indicate real change (i.e., beyond measurement error) at the 95% confidence level. The SDD for the PIPS is 4.6 points or 5.6% of the possible score range of the PIPS, indicating that two assessments of the same child should differ by >5 points to reflect a factual imitation change.

Systematic errors can be attributed to the result of learning effect. Imitation is a strong learning strategy. To avoid this error, we offered two equivalent halves of the PIPS with a 1-wk interval. Test–retest reliability was...
excellent. The small variability in the performance over the time span of 1 wk indicates that the score is highly dependent on the children’s aptitude to imitate and that this skill is a stable phenomenon, at least in a short-term period in typically developing children.

This study has some implications for clinical practice. Standardized cognitive and motor tests for young children are administered by the examiner demonstrating the tasks. Problems with these tests are the imitative demands inherent in their instructions. For that reason, we suggest that cognitive and motor abilities need to be assessed in conjunction with a formal assessment of imitation abilities. This imitation assessment may provide insight into the future learning capacity of young children and, accordingly, their possible response to intervention.

Directions for further research include (1) the replication of the findings in other samples of preschoolers, including those with atypical development and those from other cultures, and (2) the investigation of the imitation development of preschoolers with intellectual disabilities associated with unimpaired imitation (e.g., Down syndrome).

Conclusion

Because imitation is a primary learning strategy of young children, it is essential that pediatric clinicians and researchers, including occupational therapists, have a reliable measure of imitation aptitude. The PIPS meets the required standards of objectivity and stability. The PIPS score was reliable when rated twice by the same rater and when rated by different raters. In addition, the small SDD value indicates that the PIPS can be used to monitor children’s improvement by different raters. The imitation score is stable over time. Further studies are needed to extend the PIPS, reliability and validity to other populations that are prone to have advanced imitation skills or imitation difficulties.

Acknowledgments

This study was part of the doctoral dissertation of Marleen Vanvuchelen and was supported by the Foundation Marguerite-Marie Delacroix (Tienen, Belgium). We thank all the children who participated as well as the following students: Nicole Hermans and Kenneth Willems (PHL University College, Belgium) and Kim Abrams and Lien Drent (Department of Rehabilitation Sciences, Katholieke Universitaet Leuven, Belgium).

References


