
Objectivity and Stability of the Preschool Imitation and Praxis Scale

Marleen Vanvuchelen, Herbert Roeyers, Willy De Weerd

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.

Vanvuchelen, M., Roeyers, H., & De Weerd, W. (2011). Objectivity and stability of the Preschool Imitation and Praxis Scale. *American Journal of Occupational Therapy, 65*, 569–577. doi: 10.5014/ajot.2010.ajot00000414

Marleen Vanvuchelen, PhD, is Professor, Department of Health Care, PHL University College, REVAL, Agoralaan Building A, B-3590 Diepenbeek, Belgium; Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium; and Professor, Department of Physiotherapy, Vrije Universiteit Brussel, Brussels, Belgium; MVanvuchelen@mail.phl.be

Herbert Roeyers, PhD, is Professor, Research Group Developmental Disorders, Ghent University, Ghent, Belgium.

Willy De Weerd, PhD, is Professor, Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium.

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 Weerd, 2007a, 2007b, 2011b, 2011c; Williams, Whiten, & 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 1. Descriptive Summary of Current Imitation Measures for Preschool Children

Supporting Reference	Measure	Age Range	Theoretical Framework	Reliability	Age-Specific Tasks for Preschoolers	Construct Validity	Action Types	Tasks (No. of Tasks)	Standardization (Sample Size)
Ayres (1989)	Sensory Integration and Praxis Test Subsets: Postural Praxis, Sequencing Praxis, Oral Praxis	4 yr–8 yr, 11 mo	Sensory Integration Theory	Interrater Test–retest	No	One domain	Bodily	si-NM-postural (17) sq-NM-gestural (9) Oral (19)	2,000
Bergès & Lézine (1963)	<i>Test d'Imitation de Gestes</i>	3 yr–6 yr, 11 mo	Neuropsychological theory	Interrater Test–retest	Yes	One domain	Bodily	si- and bi-NM-gestural (36)	489
Korkman, Kirk, & Kemp (1998)	Developmental Neuropsychological Assessment Subset: Sensorimotor-Imitating Hand Postures	3 yr–4 yr, 11 mo 5 yr–12 yr, 11 mo	Assessment of adults with acquired brain damage	Interrater Test–retest Split-half	Yes	One domain	Bodily	si-NM-gestural (5)	1,000
Uzgiris & Hunt (1987)	Uzgiris–Hunt Scales of Psychological Development Subset: Gestural Imitation	4 mo–1 yr, 8 mo	Cognitive Developmental Theory of Piaget	Interrater Test–retest	No	No	Bodily Procedural	Familiar (4–6) Unfamiliar (4–6)	None

Note. NM = nonmeaningful, si = single; sq = sequential; bi = bimanual.

(single and sequential), and visual monitoring possibilities (transparent and opaque). Imitation tasks that may be performed by young children but are unlikely to be exhibited spontaneously were selected (Vanvuchelen et al., 2011a). Imitation age-equivalent scores are derived from PIPS scores of 654 typically developing children ages 12–59 mo (Vanvuchelen, 2009).

This article provides evidence of rater and test–retest reliability of the PIPS in typically developing children. Establishment of the objectivity or minimal measurement errors during the collection of data is critical. Intra- and interrater reliability indicate an assessment’s relative reliability. Because correlations are highly influenced by the range of values in the sample, it is also important to calculate the absolute reliability (Atkinson & Nevill, 1998), including the limits of agreement (Bland & Altman, 1986), the standard error of measurement (Atkinson & Nevill, 1998), and the smallest detectable difference (Kropmans, Dijkstra, Stegenga, Stewart, & de Bont, 1999; van Baalen, Odding, van Woensel, & Roebroek, 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

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

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; SQB = 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 4 is given if the child has used a symbolic

grip and has performed a repetitive action on both sides of the head. Score 3 is given if the child has used a symbolic grip and has performed a repetitive action on one side of the head or a single action. Score 2 is given if the child has used a body-part-as-an-object grip and has performed a repetitive action on both sides of the head. 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: $\geq 70\%$ 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 $\pm 2 SD$) 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_{first} and SEM_{second}) 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 $\pm 1.96 \times SEM$. 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, Oding, van Woensel, & Roebroek, 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):

$$SDD = 1.96 \times \sqrt{SEM_{\text{first}}^2 + SEM_{\text{second}}^2}.$$

If the difference between both $SEMs$ was <1%, we assumed that they were equal and simplified the formula as follows:

$$SDD = 1.96 \times SEM \times \sqrt{2}$$

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: 64%, t-MG3 and t-MG 4, and 85%, sq-NMG1. The interrater 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 be-

Table 3. Weighted κ Values and Percentages of Agreement for Intra- and Interrater Agreement of 30 Individual Item Scores of the Preschool Imitation and Praxis Scale

Item No.	Task	Intrarater Reliability, $N = 21$			Interrater Reliability, $N = 42$		
		κ	ASE	AG%	κ	ASE	AG%
1	sao-P1	1.00	0.00	100	0.91	0.06	95
2	sao-P2	1.00	0.00	100	1.00	0.00	100
3	sao-P3	1.00	0.00	100	1.00	0.00	100
4	i-MG1	0.84	0.09	90	0.88	0.05	90
5	i-MG2	0.92	0.07	95	0.75	0.07	78
6	i-MG3	0.93	0.06	95	0.88	0.05	90
7	si-NMG1	0.82	0.08	80	0.70	0.08	76
8	si-NMG2	0.82	0.08	85	0.78	0.06	78
9	si-NMG3	0.80	0.08	80	0.89	0.04	88
10	sq-P1	0.83	0.07	80	0.85	0.04	80
11	sq-P2	0.57	0.15	71	0.74	0.07	73
12	sq-P3	0.69	0.14	80	0.69	0.08	76
13	t-MG1	0.68	0.11	71	0.75	0.07	76
14	t-MG2	—	—	76	—	—	64
15	t-MG3	—	—	90	—	—	64
16	bi-NMG1	0.78	0.10	80	0.79	0.06	76
17	bi-NMG2	0.84	0.08	85	0.87	0.05	88
18	bi-NMG3	0.51	0.14	61	0.64	0.07	61
19	fa-NMG1	0.81	0.08	85	0.74	0.08	76
20	fa-NMG2	0.83	0.08	85	0.82	0.06	83
21	fa-NMG3	0.74	0.10	76	0.90	0.04	88
22	aso-P1	1.00	0.00	100	0.75	0.08	78
23	aso-P2	0.92	0.07	95	1.00	0.00	100
24	aso-P3	1.00	0.00	100	1.00	0.00	100
25	sq-NMG1	—	—	57	—	—	85
26	sq-NMG2	—	—	61	0.74	0.09	88
27	sq-NMG3	0.90	0.07	90	0.87	0.06	90
28	f1	0.73	0.13	80	0.88	0.05	90
29	f2	0.81	0.10	85	0.66	0.08	69
30	f3	0.86	0.08	90	0.76	0.07	78

Note. — = κ statistics could not be computed because of the skewed distribution of the data. κ = weighted κ value; ASE = asymptotic standard error; AG% = percentage of agreement; 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.

tween 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 ($r = .93$, $p < .001$).

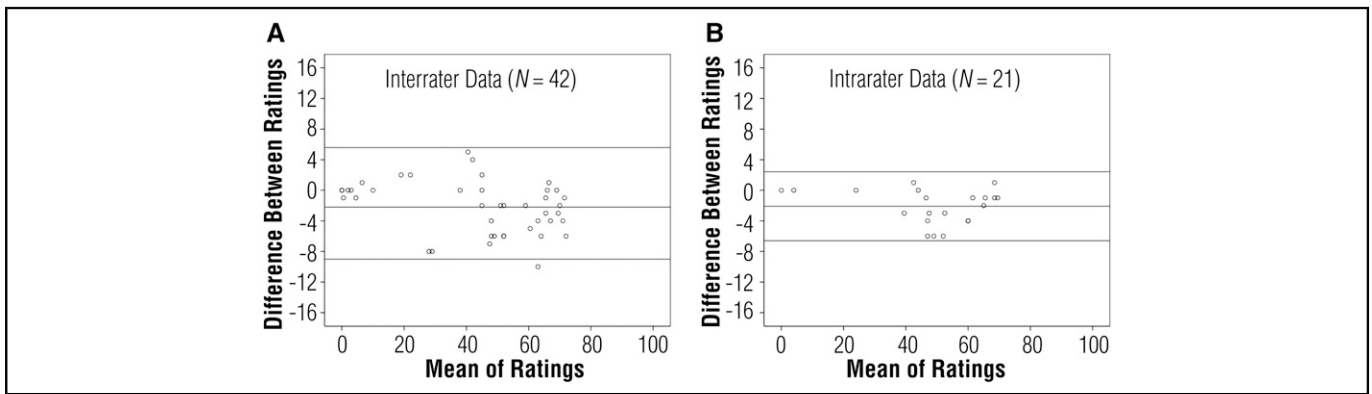


Figure 1. Bland and Altman's (1986) limits of agreement for the Preschool Imitation and Praxis Scale score on the intrarater and interrater data.

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

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant

- to sports medicine. *Sports Medicine*, 26, 217–238. doi: 10.2165/00007256-199826040-00002
- Ayres, J. A. (1989). *Sensory Integration and Praxis Test (SIPT)*. Los Angeles: Western Psychological Services.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio, TX: Harcourt Assessment.
- Bergès, J., & Lézine, I. (1963). *Test d'imitation de gestes. Techniques d'explorations du schéma corporel et des praxis chez l'enfant de 3 à 6 ans* [The gestural imitation test. A method to explore body awareness and praxis in children between 3 and 6 years of age]. Paris: Masson.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307–310.
- Bricker, D., & Squires, J. (1999). *Ages and Stages Questionnaires (ASQ): A parent-completed, child-monitoring system* (2nd ed.). Baltimore: Brookes.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- Fleiss, J. L. (1981). *Statistical methods for raters and proportions*. New York: Wiley.
- Folio, M. R., & Fewell, R. R. (2000). *Peabody Developmental Motor Scales* (2nd ed.). Austin, TX: Pro-Ed.
- Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: A comparison with specific developmental disorder of motor function. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 43, 655–668. doi: 10.1111/1469-7610.00054
- Hurley, S., & Chater, N. (2005). Introduction: The importance of imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation. From neuroscience to social science. Vol. 2: Imitation, human development, and culture* (pp. 1–52). Cambridge, MA: MIT Press.
- Korkman, M., Kirk, U., & Kemp, S. (1997). *NEPSY—Developmental neuropsychological assessment*. San Antonio, TX: Psychological Corp.
- Kropmans, T. J., Dijkstra, P. U., Stegenga, B., Stewart, R., & de Bont, L. G. (1999). Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. *Journal of Dental Research*, 78, 784–789. doi: 10.1177/00220345990780031101
- Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences USA*, 104, 19751–19756. Retrieved March 27, 2011, from www.ncbi.nlm.nih.gov/pmc/articles/PMC2148370/bin/pnas.0704452104_index.html doi: 10.1073/pnas.0704452104
- Macedoni-Luksic, M., Greiss-Hess, L., Rogers, S. J., Gosar, D., Lemons-Chitwood, K., & Hagerman, R. (2009). Imitation in fragile X syndrome. Implications for autism. *Autism*, 13, 599–611. doi: 10.1177/1362361309337850
- Masur, E. F. (2006). Vocal and action imitation by infants and toddlers during dyadic interactions: Development, causes and consequences. In S. J. Rogers & J. H. G. Williams (Eds.), *Imitation and the social mind. Autism and typical development* (pp. 27–47). New York: Guilford Press.
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

- Petreska, B., Adriani, M., Blanke, O., & Billard, A. G. (2007). Apraxia: a review. *Progress in Brain Research*, *164*, 61–83.
- Prinz, W. (2002). Experimental approaches to imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 143–162). Cambridge, England: Cambridge University Press.
- Rogers, S. J., & Williams, J. H. G. (2006). *Imitation and the social mind: Autism and typical development*. New York: Guilford Press.
- Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine*, *21*, 3431–3446. doi: 10.1002/sim.1253
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.
- Uğiris, I. C., & Hunt, J. M. (1987). *Infant performance and experience: New findings with the ordinal scales*. Urbana: University of Illinois Press.
- van Baalen, B., Odding, E., van Woensel, M. P., & Roebroek, M. E. (2006). Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. *Clinical Rehabilitation*, *20*(8), 686–700.
- Vanvuchelen, M. (2009). *Imitation problems in children with autism spectrum disorders. A study of their nature, clinical significance and utility in diagnosis*. Unpublished doctoral dissertation, Katholieke Universiteit Leuven, Belgium.
- Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2007a). Nature of motor imitation problems in school-aged boys with autism: A motor or a cognitive problem. *Autism*, *11*, 225–240. doi: 10.1177/1362361307076846
- Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2007b). Nature of motor imitation problems in school-aged males with autism: How congruent are the error types. *Developmental Medicine and Child Neurology*, *49*, 6–12. doi: 10.1017/S0012162207000047.x
- Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2011a). Development and initial validation of the Preschool Imitation and Praxis Scale (PIPS). *Research in Autism Spectrum Disorders*, *5*(1), 463–473. doi: 10.1016/j.rasd.2010.06.010
- Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2011b). Do imitation problems reflect a core characteristic in autism? Evidence from a literature review. *Research in Autism Spectrum Disorders*, *5*(1), 89–95. doi: 10.1016/j.rasd.2010.07.010
- Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2011c). Imitation assessment and its utility to the diagnosis of autism: Evidence from consecutively clinical preschool referrals for suspected autism. *Journal of Autism and Developmental Disorders*, *41*(4), 484–496. doi: 10.1007/s10803-010-1074-z
- Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*, 285–299. doi: 10.1023/B:JADD.0000029551.56735.3a