Manuscript Details

Manuscript number	HMS_2019_858
Title	A new protocol for assessing action observation and imitation abilities in children with Developmental Coordination Disorder: a feasibility and reliability study.
Article type	Registered Report Stage 1: Proposal

Abstract

AIMS. To develop a new protocol for the assessment of action observation (AO) abilities and imitation of meaningful and non-meaningful gestures, to examine its psychometric properties in children with DCD and typically developing (TD) children. BACKGROUND. For learning manual skills, AO and imitation are considered fundamental abilities. Knowledge about these modalities in children with DCD is scarce and an assessment protocol is lacking. METHOD. The protocol consists of 2 tests. The AO test consists of two assembly tasks. The imitation test includes 12 meaningful and 20 non-meaningful gestures. Items of both tests are rated on a 4-point scale. Twelve children with DCD (mean age 8y3m, SD, 1.30) and 11 TD children (mean age 8y2m, SD 1.52) were enrolled. For inter-rater reliability, intraclass correlation coefficients (ICC) were calculated for the total score, weighted kappa and percentage agreement for single items. Known group validity was assessed by comparison of DCD and TD group (Wilcoxon rank sum test). For construct validity, the mABC-2 test was used. The protocol was adapted and confirmed by an intra and inter-rater reliability study (new sample of 11 DCD children, mean age 7y5m, SD 1.37). RESULTS. Excellent ICCs were reported for intra and inter-rater reliability for the final protocol. A significant difference between DCD and TD group was found for AO abilities (p<.01), for non-meaningful gestures (p<.001). A significant correlation was reported between the AO test and the mABC-2 test (r=56;p=<.0001). No significant correlations were revealed for the imitation tests. DISCUSSION AND CONCLUSION. The results support the psychometric properties of this protocol. When fully validated, it may contribute to map the deficits in AO abilities and imitation, to evaluate treatment effects of imitation and AO interventions.

Keywords	Keywords: action observation, imitation, meaningful and non-meaningful gestures, developmental coordination disorders

Corresponding Author	Eleonora Bieber
Order of Authors	Eleonora Bieber, Bouwien Smits-Engelsman, Giuseppina Sgandurra, Francesca Di Gregorio, Andrea Guzzetta, Giovanni Cioni, Hilde Feys, Katrijn Klingels
Suggested reviewers	Jessica Lust, Deborah Dewey

Submission Files Included in this PDF

File Name [File Type]

Paper.doc [Manuscript File]

Figure 1.docx [Figure]

Figure 2.docx [Figure]

Figure 3.docx [Figure]

Figure 4.doc [Figure]

Figure 5.doc [Figure]

Appendix A.doc [Figure]

Table 1.docx [Table]

Table 2.doc [Table]

Table 3.docx [Table]

Table 4.doc [Table]

Table 5.docx [Table]

Table 6.doc [Table]

1.INTRODUCTION

Developmental Coordination Disorder (DCD) is a clinical condition characterized by reduced motor coordination and low abilities to learn motor skills at an age-appropriate level. DCD affects about 6% of school-aged children (Reynolds et al., 2015). Children with DCD experience problems in the performance of manual skills tapping into various domains of the International Classification of Functioning and Disability-Children Youth Version (ICF-CY)(Blank, 2012). At body function level, they show lower strength levels, slowness, and deficits in visual proprioceptive and kinesthetic perceptions (Raynor, 2001) (Schoemaker et al., 2001). At activity level, difficulties in execution of fine motor skills, such as in-hand manipulation, bimanual coordination (Raynor, 2001), target-direct reaching and handwriting skills have been reported (Smits-Engelsman, Niemeijer, & Van Galen, 2001) (Huau, Velay, & Jover, 2015); (Feder & Majnemer, 2007). Moreover, children with DCD may also face problems with motor planning (Adams, Lust, Wilson, & Steenbergen, 2014) (Noten, Wilson, Ruddock, & Steenbergen, 2014). All these problems may impact their academic progress, social integration and emotional development (Summers, Larkin, & Dewey, 2008). Children with DCD have difficulties learning to perform new manual skills in daily life, which typically developing (TD) peers seem to acquire almost effortlessly (Schoemaker et al., 2001). For learning and consolidating manual skills, *action observation* and *imitation* are considered fundamental capabilities, both of which are mediated by the mirror neuron system (Reynolds et al., 2017). This fronto-parietal network has an integrative role in these processes, firing when a person observes and imitates actions (Jeannerod, 2001) (Filimon, Nelson, Hagler, & Sereno, 2007) (Fabbri-Destro & Rizzolatti, 2008). Action observation and imitation activates the neural circuit partially overlapping with the circuit activated by movement execution. The existence of this shared neural circuit supports the hypothesis that action observation and imitation can thus enhance motor learning.

Action observation refers to the ability to observe a goal-directed action with the intent to imitate it. Although we hypothesize that action observation plays an important role in learning motor skills, this has not yet been studied in children with DCD and in typically developing children. In neurological populations such as adults with stroke (Pomeroy et al., 2005);(Garrison, Winstein, & Aziz-Zadeh, 2010), Parkinson's disease (Caligiore, Mustile, Spalletta, & Baldassarre, 2017) and children with unilateral cerebral palsy (Sgandurra et al., 2011), action observation training has been shown to lead to significant improvement of motor function and increased neural activity in the mirror neuron system. However, thus far, no studies measured the deficits in action observation abilities in children with DCD and a reliable and standardized assessment is presently lacking.

At a behavioral level, deficits in *imitation* have been reported in children with DCD (Reynolds et al., 2015); (Zoia, Pelamatti, Cuttini, Casotto, & Scabar, 2002) and may underlie the motor difficulties characteristic for DCD (Werner, 2012). Imitation can be defined as "the capacity to acquire, by observation, a new motor behavior and to repeat it using the same movement used by demonstrator"(Call J, 2005). Gestures represent a modality to assess imitation abilities. In the imitation of gestures, a distinction can be made between meaningful (representational) and non-meaningful gestures. Meaningful gestures are usually further divided into transitive gestures, which require the use of an object (e.g. using a toothbrush), and non-transitive gestures (e.g. wave goodbye)(Sinani, Sugden, & Hill, 2011). Non-meaningful gestures are represented by positions (novel gestures) and sequences (consecutive gestures).

As reported by Reynolds (Reynolds et al., 2015), different studies have investigated imitation abilities in children with DCD using tests of meaningful gestures). These studies indicated that children with DCD have poorer imitative gestural performance when compared to age-matched controls (Dewey, 1993); (Kaplan, 2009) (Hill, 1998);(Sinani et al., 2011);(Zoia et al., 2002) and perform similarly to younger control children(Hill, 1998);(Sinani et al., 2011). For non-meaningful gestures, far less research has been performed with inconsistent results. A few studies indicated that children with DCD are significantly less capable in imitating body postures and sequences and make more errors as tasks become more difficult compared to control groups (Reynolds et al., 2017);(Ruttanathantong, Siritaratiwat, Sriphetcharawut, Emasithi, Saengsuwan, 2013); (Kaplan, 2009). On the contrary, in Hill's study (1998) no significant difference between groups was revealed for the imitation of non-meaningful gestures. Previous studies also have some limitations in assessing non-meaningful imitation. Although tests for overall praxis (Kaplan, 2009); (Vanvuchelen, Roeyers, & De Weerdt, 2007);(Dobigny-Roman, Dieudonne-moinet, Tortrat, Verny, 1998) show strong psychometric properties in general, this has not been reported specifically for non-meaningful gestures. They often used general assessments of imitation including also for example oral praxis. Currently, no standardized and reliable assessment is available specifically for imitation of non-meaningful gestures for hand and finger positions and sequences in children.

In summary, knowledge on action observation and imitation deficits in children with DCD is scarce. To attain a better understanding of the problems in motor learning in children with DCD, we need to map the possible deficits of underlying mechanisms such as action observation and imitation.

Therefore, a first aim of this study was to develop a comprehensive assessment protocol investigating action observation and imitation proficiency for typically developing children and children with DCD.

This novel protocol will aid in mapping the deficits in action observation and imitation abilities, which in turn may facilitate delineating individualized rehabilitation programs in children with DCD focusing on action observation and imitation.

Secondly, we examined the intra and inter observer reliability, known group validity and construct validity of this protocol in typically developing children and children with DCD.

2.METHOD

The method section first outlines the developmental process of a protocol for assessing action observation and imitation abilities in children with DCD and typically developing children aged 6-10 years old. The protocol includes three tests: a test for action observation abilities, a test for non-meaningful gestures and a test for meaningful gestures. Secondly, the methodology of investigating psychometric properties of the protocol is reported. We performed a first study on reliability, known group and construct validity on the designed protocol in DCD and TD children. Based on the administration and expert panel discussion, final adaptations in the material and formulation of the scoring system were made in order to reduce the length of administration and to improve the test standardization. We then performed a second study in children with DCD to confirm the reliability.

The developmental process and investigation of psychometric properties of the protocol are visualized in the flowchart (Figure 1).

2.1. Test for action observation abilities

Since no test for action observation existed, we developed a new test that will be described in the next section.

2.1.1. Background

Critical literature review revealed no existing assessments of action observation abilities in children or adults. Therefore, the development of a new test was based on insights on activation of the mirror neuron system. First, if observation is accompanied by imitation, a higher activation of the mirror neuron system is expected (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005) (Cross, Hamilton, & Grafton, 2006). Secondly, it is known that mirror neurons discharge in association with observation and imitation of specific motor acts (e.g. grasping, holding, tearing) rather than simple movements (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Thirdly, the mirror neuron system is more strongly involved in goal-directed and novel actions (Vogt & Thomaschke, 2007) (Sgandurra et al., 2011).

Based on expert panel discussion and a piloting phase, we selected a simple and complex assembly task, using wooden and metal material according to 6-7 years and 8-10 years old children, respectively. We further modified the scoring system of Watkins (Watkins, Dronkers, & Vargha-Khadem, 2002) and reformulated the scoring criteria. Based on the experience of test administration during the first reliability and validity study and expert panel discussion, we opted to use only wooden material in order to improve the standardization of the test.

2.1.2. Description of the test

The children are asked to perform two assembly tasks, a simple task (screw a bolt on a nut) and a complex task (compile a triangle with three sticks and three nuts and bolts, Figure 2). First, the child has to perform the simple task spontaneously in order to see which strategy he uses. Secondly, the child observes a video clip on a laptop screen showing a strategy (different phases that represent the action) for the simple task and is subsequently asked to imitate the strategy. Two trials of video observation and imitation are performed. Finally, a second strategy for the simple task is shown which the child has to observe and imitate. Again, two trials are performed. For the complex task, after the spontaneous execution, only one strategy is observed. We limited the imitation to one trial for the complex task, because of the long duration of the task. If the performance of the child lasts more than two or four minutes, for the simple and complex task respectively, the examiner interrupts the execution and moves to the next trial. Video clips for simple and complex tasks are shown from the first person perspective and are adapted according to hand dominance. During the test administration, the examiner is sitting next to the child ensuring the focus of attention on the laptop screen. The performance of the tasks is video recorded and time-watched. The test administration lasts 15 minutes and scoring takes 20 minutes.

2.1.3. Scoring system

The AO performance is scored according to a four-point ordinal scale, with a score of 0 for a fully incorrect imitation of the strategy or not able to perform the task; a score of 1 for an attempt at the correct imitation of the strategy; a score of 2 for a partially correct imitation of the strategy but problem in execution; and a score of 3 for a correct imitation with good execution. Also, the time of the total task is registered.

2.2. Tests for imitation abilities

Imitation abilities are investigated by the administration of non-meaningful gestures (development of a new test) and meaningful gestures (use of an existing test).

2.2.1. Non-meaningful gestures

2.2.1.1. Background

Although imitation of novel actions is likely of the greatest relevance with reference to the mirror neuron system, imitation of non-meaningful gestures and sequences is hardly included in the assessment of praxis ability. Literature review revealed three studies assessing non-meaningful gestures in children (Goyen, Lui, & Hummell, 2011) (Dewey, Kaplan, Dewey, & Kaplan, 1992); (Hill, 1998). Only three out of six subtests of the assessment battery proposed by Goyen (Goyen et al., 2011) incorporated non-meaningful gesture imitation: the postural praxis (body movements and position), oral praxis (lip, tongue and jaw movements) and sequencing praxis (series of hand and finger movements). Limitations to existing non-meaningful gesture tests are the small number of gestures and their simplicity and the lack of studies on psychometric properties (Reynolds et al., 2015). For our specific interest in manual skills, we therefore opted to develop a new test focusing on hand and finger positions and sequences, hereby including a larger number of items ranging from unimanual to bimanual, from simple to complex positions as well as sequences. Based on expert discussion, we modified the original scoring system of Watkins et al. (2002), already used in other experimental gestures protocol in children with autism spectrum disorder (Cossu et al., 2012). Based on the experience of test administration during the first reliability and validity study and expert panel discussion, we opted to show the children video-clips from the third perspective rather than real life presentation of the examiner to facilitate the administration procedure and standardize the procedure.

2.2.1.2. Description of the test

The test of non-meaningful gestures and sequences consists of eight hand and eight finger positions and four sequences with one or two hands (Appendix A). After video observation, the child imitates the observed gesture or sequence. Three different conditions are used: 'delayed imitation', 'concurrent imitation', and 'explanation condition'. In the 'delayed imitation', the video clip shows the gesture for 5 seconds. The child is asked to reproduce the observed gesture immediately afterwards. Subsequently, the second trial or 'concurrent imitation' is performed, in which the gesture is shown for 20 seconds and the child can start to reproduce the gesture while observing. If the child is able to imitate the gesture in the second trial, the examiner moves on to the next gesture. If the child is unable to imitate the gesture, a third trial is performed. In the third trial or 'explanation condition', the video clip presents the movement sequences of the gesture. The child is asked to imitate the gesture while observing the movement sequences. The imitation performance of the children is video recorded. The test administration takes 10 minutes and 20 minutes for scoring.

2.2.1.3. Scoring system

The imitation performance is scored according to a four-point ordinal scale (modification of Watkins et al., 2002). A score of 0 is for no or incorrect execution in any of the conditions; score of 1 for a correct execution in the explanation condition; a score of 2 for a correct execution in the concurrent imitation condition and a score of 3 for a correct execution in the delayed imitation condition. The imitation of the hand and finger position is rated as 'correct execution' when: (1) the hand and finger orientation is equal to the demonstration, and (2) the performance is executed within 5 seconds from the demonstration. The imitation of the sequence is rated as 'correct execution' if it is performed with an acceptable quality at least one time without any movement interruption. Total score ranges from 0 to a maximum score of 60.

2.2.2. Meaningful gestures

2.2.2.1. Background

Meaningful gestures represent a type of imitation ability which has typically been assessed in children with DCD, for example in the Sensory and Integration Praxis Test (Goyen et al., 2011) or the Representational Gestures Praxis test, (Hill, 1998) (Sinani et al., 2011)(Zoia et al., 2002)(Dewey et al., 1992). Also, in other populations such as children with autism spectrum disorder and learning disorders, different imitation assessments have been used including the ideomotor and praxis abilities test (Glosser & Wiener, 1986) (De Renzi, 1985) (Rothi, Ochipa, Heilman, 1997) the ideomotor Apraxia Test (Dobigny-Roman,Dieudonne-moinet,Tortrat, Verny, 1998) and the Preschool imitation and Praxis Scale (Vanvuchelen, Roeyers, & Weerdt, 2009).

After critical literature review for identifying tests feasible for imitation abilities in school-aged children with DCD, we selected the representational (or meaningful) gestures part of the Gestures Test developed by Dewey and colleagues (Dewey et al., 1992) (Dewey, Roy, Squarer-Storer, Hayden, 1988). This test has been used in children with developmental motor disorders, aged 6-10 years old, and has high inter-rater reliability (Dewey et al., 1992). For our protocol, we only selected the part on verbal command of the meaningful gestures test (Dewey et al., 1992). The Gesture Test consists of three praxis measures: representational (or meaningful) gestures test (imitation and verbal command), sequencing gestures test and non-representational (or non-meaningful) gestures tests.

2.2.2.2. Description of the test

The meaningful gestures part of the Gestures Test (verbal command) (Dewey et al., 1992) includes six transitive gestures (with objects) and six intransitive gestures (without objects) (Table 1).

For administration, children are seated in a comfortable position in front of the examiner and asked to show how a specific gesture is performed, e.g., "Show me how you brush your teeth with a toothbrush." Participants are videotaped, and gesture performance is scored from videotapes. The test administration lasts 5 and 15 minutes for scoring.

2.2.2.3. Scoring system

The original scoring system is used as described in the Gesture Test (Dewey et al., 1992): a score of 0 is given if an *incorrect gesture* is performed or if the children indicate *where* the gesture is to be performed (e.g., pointing to the mouth for brushing teeth); a score of 1 if the children demonstrate a *body-part-as-object* error (e.g., using the index finger extended to make brushing movements against the teeth), or an error in *orientation* (e.g., incorrect plane of movement), *posture* (e.g., incorrect hand posture), *location* (e.g., gesture performed in incorrect spatial location), or *distortion* (e.g., distortion in timing of the movement); a score of 2 if there are *minor inaccuracies* in performing the gesture (e.g., extension or helper error); and a score of 3 if the *action* is performed *correctly*. A classification of gestural errors (published in

Dewey, 1993) is provided in order to help the examiner in the scoring process. The total score ranges from 0 to 36.

2.3. Reliability and validity study

2.3.1. Participants

Children with DCD were recruited in Belgium via 'vzwDyspraxis' and by contacting local physiotherapists, and in Italy in the 'Stella Maris Institute' (Pise). Included children with DCD all met the four DSM-V diagnostic criteria for DCD (American Psychiatric Association, 2013); (1) "coordinated motor skills substantially below that expected given the individual's chronological age and opportunity for skill learning and use", as evidenced by mABC-2 total percentile score ≤ 16^{th} (Henderson E, et al., 2007), (2) "motor skills deficit significantly and persistently interferes with activities of daily living", as evidenced by the DCD Daily (Van der Linde et al., 2013), DCDQ'07 questionnaire (Parmar, Kwan, Rodrigues, Missiuna & Cairney, 2014; Wilson & Crawford, 2012) and mABC-2 Checklist (Henderson et al., 2007; Schoemaker, Niemeijer, Flapper and Smits-Engelsman, 2012) (3) "onset of symptoms is in the early developmental period" as evidenced by anamnestic information and pediatrician consultant referee; (4) "motor skills deficits not better explained by intellectual disability or visual impairment and not attributable to a neurological condition", as evidenced by psychological and neurological assessment.

Age matched TD children were recruited via colleagues, friends and families of the investigators both in Belgium and Italy. Children in the TD group met the following criteria: (1) score above percentile 25 on the mABC-2, (2) spoke and understood Flemish or Italian and (3) were sufficiently cooperative when performing the tests.

For the first reliability and validity study, the protocol was tested in a sample of 12 children with DCD (8 boys and 4 girls) aged between 6 and 10 (mean age 8y3m, SD 1.30), and in 11 aged matched (mean age 8y2m, SD1.52) typically developing children (6 boys and 5 girls). The population of the confirmatory reliability study consisted of a sample of 11 children (9 boys and 2 girls) with DCD (mean age 7y5m, SD 1.37), recruited in the 'Stella Maris Scientific Institute', Pisa (Italy). Participants met the same inclusion and exclusion criteria of the first reliability and validity study.

The studies were approved by the Ethics Committee of UZ KU Leuven/research 3/3/2016, s58819 and by the Ethical Committee (43/2017) of Department of Developmental Neuroscience of IRCCS Stella Maris (Pisa, Italy). All parents provided written informed consent and children above 7 years old written assent.

2.3.2. Procedure

For the first reliability and validity study, two neuropediatric therapists (EB, KK) performed the tests of the imitation and action observation protocol. Children performed all the tests in the same session in a quiet room. Total test time was around 30 minutes. Afterwards, two neuropediatric therapists (EB, FDG) independently scored the videos for inter-rater reliability after they had been trained for the scoring system using videoclip's examples. For intra-rater reliability, EB scored all the videos a second time after 14 days. For the second reliability study, a neuropediatric therapist (EB) administered the test of the imitation and action observation. To examine inter-rater reliability, two neuropediatric therapists (EB, FDG) independently scored the videos. For intra-rater reliability, EB scored all the videos a second time after 14 days.

2.3.3. Statistical analysis

For both reliability studies, intra-rater and inter-rater reliability for the total scores of the tests were determined by intraclass correlation coefficients (ICC, two-way random effect model). The 95% confidence intervals (CI) were reported. ICC values > 0.90 were considered excellent, 0.75-0.90 good and <0.75 poor to moderate (Portney & Watkins, 2009). The reliability at item level was assessed by weighted kappa values and percentage of agreement. Kappa values of>0.80 were considered excellent, 0.61–0.80 substantial, 0.41–0.60 moderate and <0.41 poor (Landis & Koch, 1977).

Known group validity was assessed by comparing the total scores of the tests of children with DCD and TD children, using the Wilcoxon rank sum test. Finally, to evaluate construct validity in children with DCD and TD children, Spearman rank correlation coefficients were calculated between the total scores of the tests (Action Observation Test; Non-Meaningful Gesture Test and the meaningful gestures part of the Gesture Test, (Dewey et al., 1992) with the standard score of the manual dexterity domain of mABC2 test and with the mABC2 test total score. For this, data from the first score of EB were used.

The level of significance was set at $p < 0.05$. Statistical analyses were conducted with the statistical software program IBM SPSS Statistic 25.

3.RESULTS

3.1. Intra and inter-rater reliability

3.1.1. Action observation test

For the DCD group, intra and inter-rater agreement of the total score showed excellent reliability, with both ICCs> 0.90 for the first and the second reliability study (table 2).

For the TD group, moderate to good intra and inter-rater reliability was found (ICC=.87 and ICC=.73 respectively).

For the DCD group, weighted kappa values for intra and inter-rater reliability were moderate to excellent (K values between 0.48 -1.0, 60%-100%) except for some conditions in the first reliability study (table 3). For the TD group, intra and inter-rater reliability ranged from moderate to excellent in all AO test conditions with K values between 0.44-1.0, 64%-100 % except for one condition.

3.1.2. Imitation tests

For the DCD group, both the first and second reliability study showed excellent reliability coefficients with ICCs >0.90 for meaningful and non-meaningful gestures tests.

ICCs and 95% CI for the total score of imitation tests for the first and second reliability study are presented in Table4. For the TD group, the first reliability study showed good agreement for the meaningful gestures test (ICCs ≥ 0.75) and excellent agreement for the Non-Meaningful Gestures Test (ICCs> 0.90).

Weighted kappa values and percentages of agreement of imitation tests for the first and second reliability study are presented in table 5 and 6.

Kappa values for individual items of the *meaningful gestures test* ranged between moderate and excellent coefficients (K between 0.42 and 1.0, 70%-100% agreement) for intra and inter-rater reliability (first and second study) in the DCD group, except for poor inter-rater agreement values for two items. In the TD group, moderate to excellent intra and inter-rater reliability coefficients were found for all items of meaningful gestures (K values between 0.46-1.0, 73%-100% agreement), except for poor values for two items.

Kappa values for individual items of *Non-Meaningful Gestures Test* in DCD ranged between moderate and excellent in both studies (K values between 0.41-1.0, 60%-100%), with the exception of one item.

For the TD group, intra and inter-rater reliability for all items of the Non-Meaningful Gestures Test ranged between substantial to excellent (K values between 0.74-1.0, 91%-100% agreement), except for poor values for inter-rater reliability of one item. Weighted kappa values and percentages of agreement of imitation tests for the first and second reliability study are presented in table 5 and 6.

3.2. Known group validity

For the total score of the action observation test, a significant difference was revealed (p=.01) between groups with a median of 58.5 (IQR 46.7-78.3) for children with DCD and 81.9 (IQR 69.4-91.7) for the TD children.

For the children with DCD, the median score for the Non-Meaningful Gestures Test was 73.3 (IQR 65.8-84.4) and 87.7 (IQR 81.6-93.3) in TD children, showing a significant difference between groups (p=.006).

For the meaningful gestures test, the median total score of children with DCD was 54.8 (44.4-61.1 IQR) and 64.3 (IQR 59.7-68.6) in TD children reaching borderline significance between groups (p=.056).

Median percentages of the total score for the three tests in children with DCD and TD children are presented in figure 3. Raw scores were converted to percentages.

3.3. *Construct validity*

A moderate correlation was found between the AO test of the total group with the total score of the mABC-2 test (r=.59, p<.0001); and a low correlation with the standard score of the manual dexterity domain of the mABC-2 test (r=.37, (p=.09) (see figure 3 and 4).

Correlation coefficients between meaningful and non-meaningful gestures tests with the mABC-2 test total score and with mABC-2 test manual dexterity standard score revealed no significant correlations (total mABC-2 score r=.01, p=.96; r= .06, p=.81; manual dexterity score r=.18, p=.44 and r=.31, p=.14 respectively).

Correlation coefficients between AO test and imitation tests were also investigated revealing no significant association between AO test total score and non-meaningful test total score (r=-.26, p=.31); and between AO total score and meaningful test total score (r=-.26, p=.30).

4. DISCUSSION

The aim of this study was to develop a novel protocol for assessing action observation and imitation abilities and to examine its psychometric properties. For this purpose we developed two new tests, one for action observation abilities and one for non-meaningful gestures, and we used the existing Gesture Test (Dewey et al., 1992) for the assessment of meaningful gestures. The results support the psychometric properties of the new tests for action observation abilities and imitation of non-meaningful gestures, and confirm reliability properties of the imitation of meaningful gestures of the Gesture Test (Dewey et al., 1992). Also, we showed that children with DCD have a reduced action observation and imitation proficiency in comparison to typically developing peers. This standardized protocol may be useful in future studies to gain more insights in action observation and imitation proficiency in children with DCD and to delineate individualized interventions, such as action observation or imitation training.

To assess action observation proficiency, we developed the *Action Observation Test*, which consists of the execution of different strategies of a simple and complex assembly task after observing videoclips showing these strategies. To the best of our knowledge, no previous study measured the deficits in action observation abilities in children with DCD and no reliable and standardized assessment is currently available in adult or pediatric populations. In the DCD group, intra and inter-rater reliability was excellent and moderate to good intra and inter-rater reliability was reported for the TD group. In both groups, weighted Kappa values ranged between moderate to excellent in all AO test conditions except for some conditions in the first reliability study. Based on the experience of the test administration during the first reliability study and in order to make the administration procedure more standardized and feasible, we opted to use wooden material for all ages, we reduced the numbers of trials for the complex tasks, and we introduced an execution time limit. Moreover, in order to facilitate the scoring process, we specifically reformulated the scoring criteria mainly focused on the imitation of the strategies rather than on the motor execution.

Imitation proficiency includes both meaningful and non-meaningful gestures. *Meaningful gestures* have previously been investigated in children with DCD, but mostly based on non-standardized assessments and on small assessment batteries (Reynolds et al., 2015). Literature search revealed the oral command of the Gesture Test (Dewey et al., 1992) as most optimal because of its feasibility for use in children with DCD and its high inter-rater reliability. Dewey et al. (Dewey et al., 1992) reported excellent inter-rater agreement for gestures on verbal command in a large sample of children with typically development (n=51) and children with developmental motor problems (n=51), aged 6 to 10 years old. Our study confirmed the excellent intra and inter-rater reliability in the DCD group and good reliability in the TD group. Based on our results, one item needs specific attention. This item required the child to show how he/she eats ice cream with a spoon. However, children mostly eat ice cream from a cone, which led to misunderstandings and imitation errors. In the second reliability study, we modified the verbal request (‘show me how you eat an ice cream in a cup’), but difficulties in the scoring were still present due to problems considering both hands. It could indeed happen that the position of one hand (the one holding the ice cream cup) was correct, but the other hand (the one holding the spoon) was not orientated in the right position. To improve the scoring standardization of this item we suggest that a specific task instruction should be added, and the score should include a carefully observation of the position of both hands.

For *Non-Meaningful gestures*, we developed a new test assessing a range of hand/finger positions and sequences, including 20 items ranging from unimanual to bimanual movements, from simple to complex positions as well as sequences. As such, this test meets the need identified by Reynolds (Reynolds et al., 2015) to develop a test with a large number of complex and novel gestures, which have shown to be related to mirror neuron system from a behavioural perspective and imitation of non-meaningful gestures and sequences in children with DCD. We found excellent intra and inter-rater reliability ICCs in both samples. Kappa values ranged from moderate to excellent values in the DCD group, and from substantial to excellent values in the TD group except for moderate inter-rater reliability on a single item. In the second study, the definition of the scoring criteria was reviewed and the gestures of the tester were replaced by videos. These adaptations further increased the reliability coefficients of almost all items, decreased the test duration and improved standardization and feasibility of the test.

To evaluate the known group validity, we compared the action observation proficiency between children with DCD and age matched controls. These results support the hypothesis that children with DCD have difficulties with action observation abilities, which might interfere with learning motor skills. Therefore, a specific intervention training based on action observation might potentially improve action observation proficiency as well as motor learning of manual skills. Action observation training has been used in neurological populations such as adults with stroke (Pomeroy et al., 2005; Garrison et al., 2010), Parkinsons's disease (Caligiore et al., 2017) and children with unilateral cerebral palsy (Sgandurra et al., 2013). In children with DCD, action observation training has previously been combined with motor imagery training (Adams, Smits-Engelsman, Lust, Wilson, & Steenbergen, 2017); (Wilson, Adams, Caeyenberghs, Thomas, Smits-Engelsman, 2016) (Reynolds et al., 2015);(Smits-Engelsman et al., 2013);(Wilson, Thomas, & Maruff, 2002). These studies showed promising results supporting the potential of action observation training to improve motor learning in children with DCD. However, the outcome of all these intervention studies was motor proficiency. Future studies should investigate the underlying mechanisms of the improvements in motor proficiency after AO training by assessing the AO abilities before and after training. AO training could be considered as a new task-oriented intervention approach for enhancing motor performance. According to scientific literature, top-down rehabilitative approaches have, in fact, been seen to show strong treatment effects (Smits-Engelsman et al., 2013).

From the analysis of known group validity for imitation proficiency, our results revealed a highly significant difference for the imitation of non-meaningful gestures in favor of the TD children and a borderline significance for the meaningful gestures, which is in accordance with previous literature (Dewey et al., 1992); (Dewey, 1993) (Sinani et al., 2011) (Zoia et al., 2002). In a group of 11 children with DCD and language disorder, (Hill, 1998) reported that they performed better on unfamiliar hand/posture sequences than on meaningful gestures. This finding was interpreted as "the familiarity of an action might lead to less reliance on visual monitoring of performance and hence less accurate execution". Also in a study of (Goyen et al., 2011) the performance of 21 preterm born children with DCD during the execution of the postural and sequencing domain of the praxis test (SIPS test) was significantly lower compared to typically developing children, with the exception of praxis on verbal command. In conclusion, this novel standardized protocol is reliable and useful for assessing differences in imitation proficiency between children with DCD and TD children. Also, the protocol seems feasible as it lasts only 15 minutes and children showed good cooperation in performing both tests (Gesture Test, Dewey et al., 1992 and the newly developed Non-Meaningful Gestures Test).

To examine the construct validity of both action observation test and imitation tests, we correlated the total scores with both the total and the subscore of the manual dexterity subdomain of the mABC2 test. A significant correlation was found with the action observation test (mainly with the total score than with the manual dexterity subdomain) revealing that lower performance on the mABC-2 test can be associated with lower action observation abilities. These results are in line with the scientific rationale that AO and motor imagery (MI) activate neural substrate partially overlapped with those activated by movement execution. For imitation we did not find significant correlations for both tests. However, further research is needed in a larger sample to disentangle the relation between learning manual skills and action observation and imitation.

This study also has some limitations. A first one is the small number of children enrolled. Future studies should consider larger sample sizes of DCD and TD children to confirm known group validity findings. Another limitation concerns the clinical comorbidity that children with DCD might have such as attention deficit, visual problems, and behavioral aspects that could interfere with the performances of imitation and action observation. In future research, specific questionnaires need to be used capturing comorbidity aspects and the impact of these comorbidities on the performance of imitation and action observation should be investigated. Still, this was the first study to investigate imitation and action observation abilities with the development of a novel and comprehensive protocol specifically developed for children with DCD. Considering the good psychometric properties, the presented protocol has the potential for use in clinical practice as well as in research to provide new information about underlying mechanisms of deficits in children with DCD,. A next step would be to map imitation and action observation abilities and their longitudinal evolution in a larger sample of children with DCD, taking the severity of the disorder and the comorbidities into account. Furthermore, another next step would be to explore the responsiveness of the imitation and action observation assessments to detect changes after specific interventions using action observation training to improve the learning of manual skills.

5. CONCLUSIONS

Action observation and imitation abilities, both mediated by the mirror neuron system, are considered fundamental abilities for learning and consolidating new manual skills. Therefore, this study presents a novel protocol, which has been shown to be reliable and feasible for the assessment of action observation and imitation abilities in children with DCD. The protocol may have great relevance to detect improvement of action observation and imitation abilities after specific intervention studies. Further studies are needed to confirm the psychometric properties in a larger sample also considering comorbidity aspects that could be present in children with DCD.

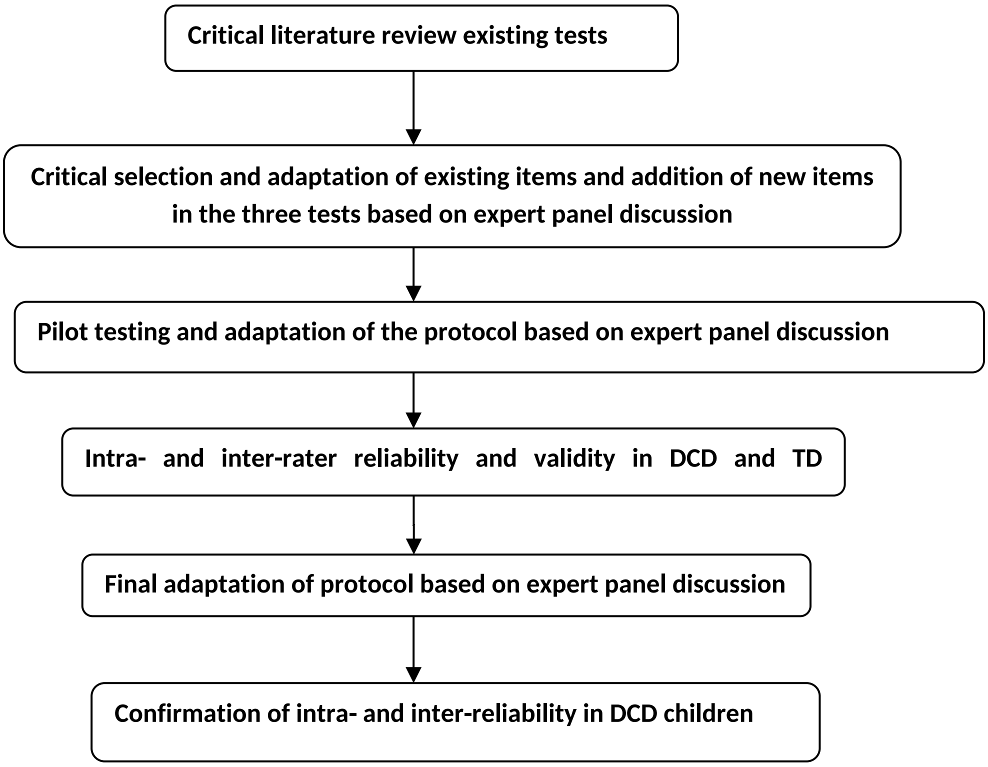


Figure 1. Flowchart of developmental process and the reliability and validity studies

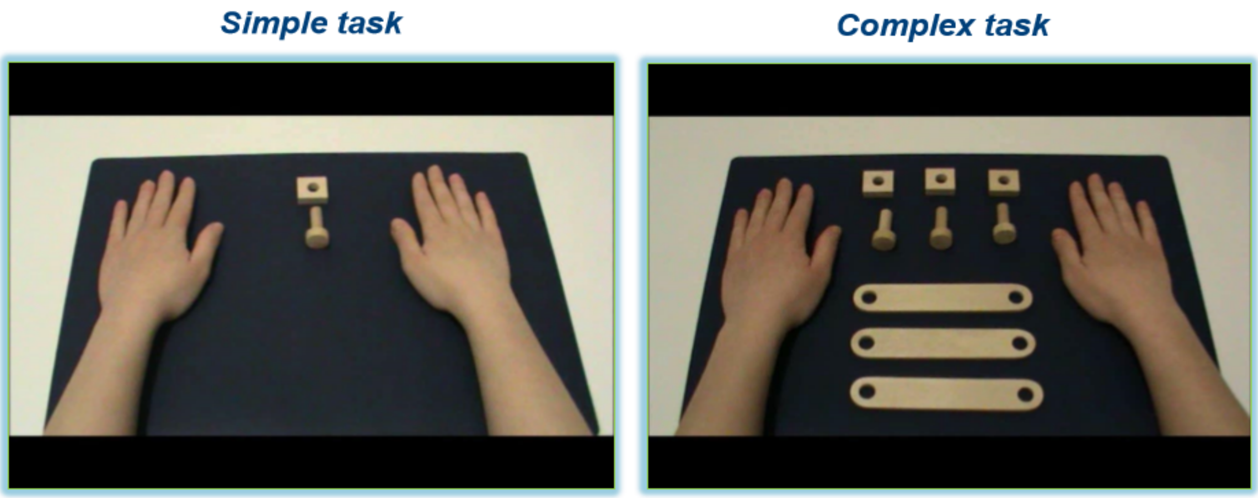


Figure 2. Action Observation test, simple and complex task

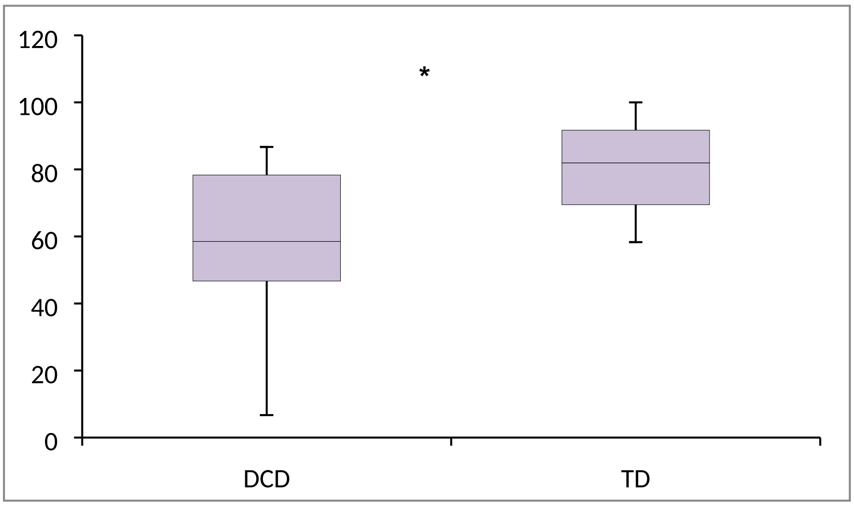


Figure 3a.

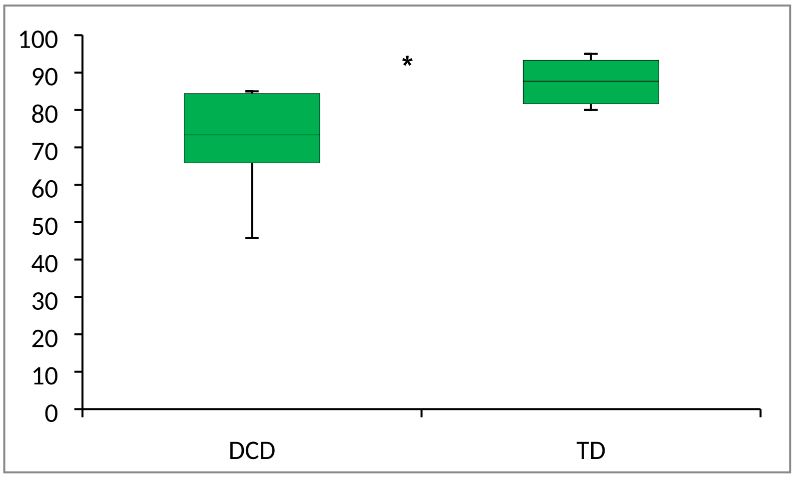


Figure 3b.

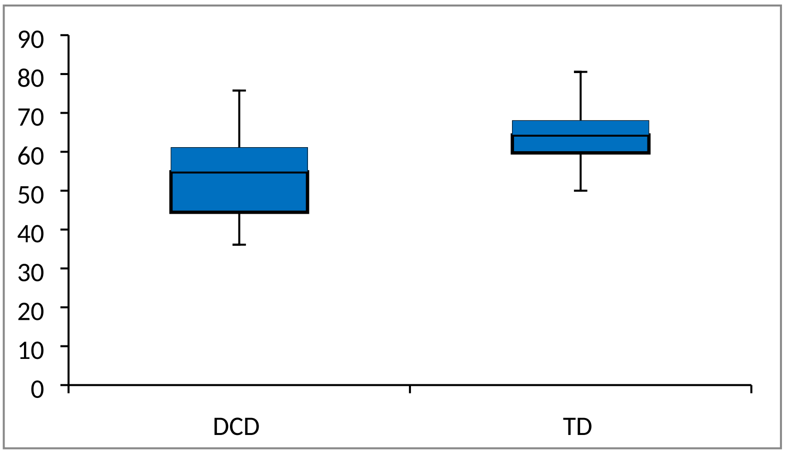


Figure 3c.

Figure 3. Box plot of difference between children with Developmental Coordination Disorder (DCD) and typically developing children (TD) in total score of and action observation test (a), Non-Meaningful Gesture Test (b), meaningful Gesture Test (c). Box plot shows: minimum, first quartile, median, third quartile and maximum.

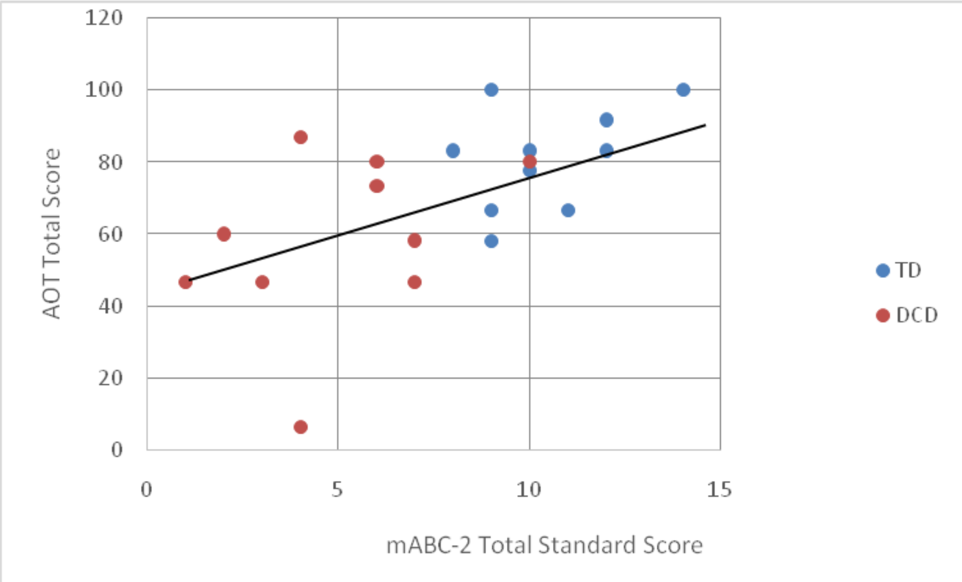


Figure 4. Scatter plot of correlation between Action Observation Test and mABC-2 total standard score in the whole sample. TD (Typically developing children); DCD (Developmental Coordination Disorder).

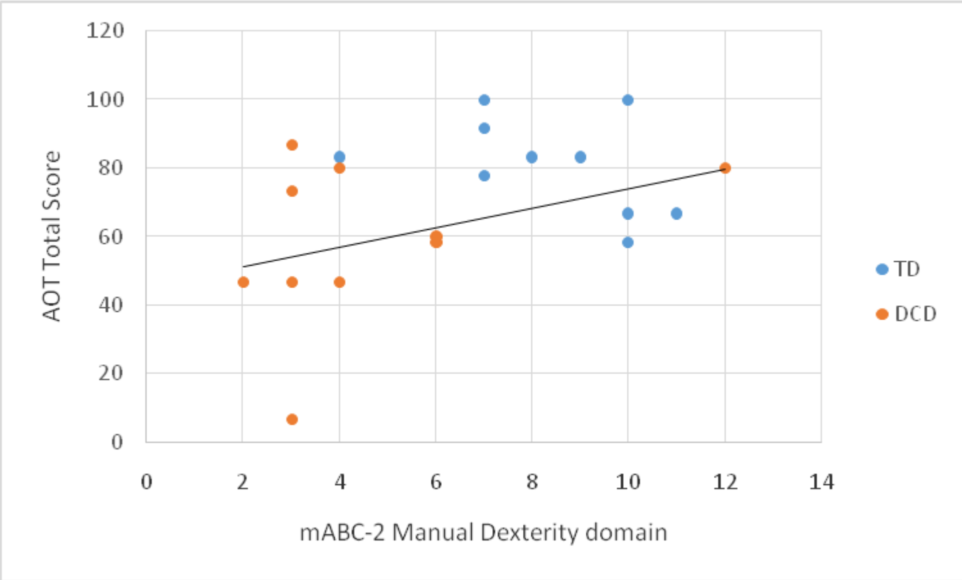
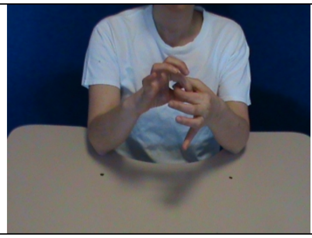


Figure 5. Scatter plot of correlation between Action Observation Test and mABC2 manual dexterity subdomain in the whole sample. TD (Typically developing children); DCD (Developmental Coordination Disorder).

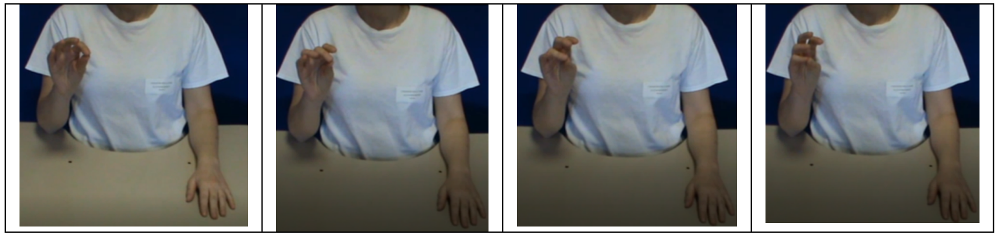
NON-MEANINGFUL ESTURES TEST

- **Hand and finger positions**

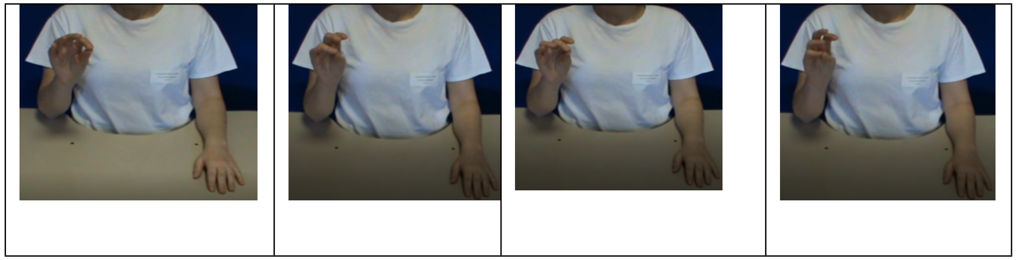
n.1	n.2	n.3	n.4
			
n.5	n.6.	n.7	n.8
			
n.9	n.10	n.11	n.12
			
n.13	n.14	n.15	n.16
			

- **Sequences**

n.17 Touch the thumb with every finger from index to little finger and back (3 times consecutively)

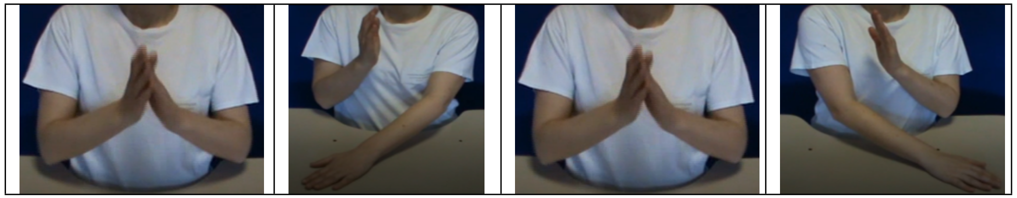


n.18 Touch the thumb with every finger in this order: index-ring finger, medium-little finger (3 times consecutively)



n.19 Clap hands in middle line, then clap both hands simultaneously on the same side of the table (5 times consecutively)

n.20 Clap hands in middle line, then cross middle line and clap one hand to the opposite side on the table (5 times consecutively)



Transitive gestures	**Intransitive gestures**
1. Brush your teeth with a toothbrush	1. Salute
2. Comb your hair with a comb	2. Pinchyournose
3. Eat ice cream with a spoon	3. Cross yourfingers
4. Hit a nail with a hammer	4. Make a fist
5. Cut paper withscissors	5. Wave goodbye
6. Write with a pencil	6. Snap your fingers

Table 1. Gesture Test (Dewey et al., 1992)

	INTRA-RATER Agreement		INTER-RATER Agreement	
GROUP	**ICC**	**95% CI**	**ICC**	**95% CI**
TD First Study	.87	.59-.96	.73	.26-.92
DCD First Study	.92	.71-.98	.91	.67-.98
DCD Second Study	.95	.84-.98	.95	.83-.98

Table 2. Intraclass correlation coefficients and 95% confidence intervals for the Action Observation Test in typically developing (TD) children and in children with DCD.

CONDITIONS	FIRST RELIABILITY STUDY								SECOND RELIABILITY STUDY			
	TD GROUP				DCD GROUP				DCD GROUP			
	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa
Simple task First trial First strategy	100%	1.0	73%	.37	80%	.66	80%	.66	80	.69	100	1.0
Simple task Second trial First Strategy	-	-	-	-	60%	.33	50%	.27	100	1.0	100	1.0
Simple task First trial Second strategy	73%	.60	64%	.44	100%	1.0	90%	.86	80	.69	80	.69
Simple task Second trial Second strategy	100%	1.0	75%	.61	60%	.48	40%	.26	80	.73	100	1.0
Complex task	82%	.65	82%	.83	78%	.70	67%	.54	90	.85	80	.71

Table 3. Weighted Kappa values and percentages of agreement for first and second study, Action Observation Test in typically developing (TD) children and in children with DCD.

-: values not available because of small sample

	MEANINGFUL GESTURES TEST				NON-MEANINGFUL GESTURES TEST			
	INTRA-RATER Agreement		INTER-RATER Agreement		INTRA-RATER Agreement		INTER-RATER Agreement	
GROUP	**ICC**	**95% CI**	**ICC**	**95% CI**	**ICC**	**95% CI**	**ICC**	**95% CI**
TD First Study	.79	.41-94	.75	.31-.92	.94	.80-.98	.93	.77-.98
DCD First Study	.95	.82-.99	.97	.88-.99	.97	.92-.99	.97	.91-.99
DCD Second Study	.99	.96-.99	.96	.77-98	.98	.95-.99	.99	.96-.99

Table 4. Intraclass correlation coefficients (ICC) and 95% confidence intervals (CI) of Non-Meaningful Gesture Test and Gesture Test for meaningful gestures in typically developing (TD) children and children with development coordination disorder (DCD), first and second reliability study

	FIRST RELIABILITY STUDY								SECOND RELIABILITY STUDY			
	TD GROUP				DCD GROUP				DCD GROUP			
	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa
1	82%	.72	82%	.85	100%	1.0	100%	1.0	90%	.83	90%	.82
2	100%	.83	73%	.50	100%	1.0	88%	.71	100%	1.0	100%	1.0
3	73%	.46	64%	.18	78%	.65	56%	.41	90%	.83	60%	.38
4	64%	.40	82%	.68	78%	.41	89%	.78	100%	1.0	100%	1.0
5	100%	.*	100%	.*	100%	.*	100%	.*	100%	1.0	100%	1.0
6	82%	.65	100%	1.0	75%	.48	100%	1.0	70%	.42	90%	.80
7	100%	1.0	80%	.69	100%	1.0	100%	1.0	100%	.83	90%	.81
8	100%	.*	100%	.*	100%	1.0	100%	1.0	100%	1.0	90%	.79
9	100%	.74	100%	.74	100%	1.0	100%	1.0	100%	1.0	80%	.70
10	82%	.67	100%	1.0	100%	1.0	100%	1.0	100%	1.0	80%	.66
11	100%	.81	82%	.62	100%	1.0	67%	.50	100%	1.0	100%	1.0
12	100%	1.0	100%	1.0	100%	1.0	100%	1.0	90%	.63	90%	.75

Table 5. Weighted Kappa values and percentages of agreement for meaningful gestures items in typically developing (TD) children and children with development coordination disorder (DCD) (Gesture Test, Dewey et al., 1992), first and second reliability study.

*= not able to calculate since the item is an unvaried variable

	FIRST RELIABILITY STUDY								SECOND RELIABILITY STUDY			
	TD GROUP				DCD GROUP				DCD GROUP			
	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa	INTRA RATER %	INTRA RATER Kappa	INTER RATER %	INTER RATER Kappa
1	91%	.*	100%	1.0	75%	.58	92%	.86	80%	.66	100%	1.0
2	91%	.84	100%	1.0	75%	.56	92%	.85	90%	.84	70%	.60
3	100%	1.0	100%	1.0	92%	.76	92%	1.0	100%	1.0	90%	.75
4	100%	.81	73%	.49	100%	1.0	83%	.75	80%	.69	70%	.52
5	100%	1.0	100%	1.0	75%	.45	75%	.45	90%	.61	100%	1.0
6	100%	1.0	91%	.81	67%	.76	100%	1.0	90%	.85	100%	1.0
7	100%	1.0	100%	1.0	100%	1.0	100%	1.0	100%	1.0	70%	.45
8	91%	.84	100%	1.0	92%	.87	92%	.88	100%	1.0	90%	.81
9	91%	.83	91%	.85	100%	1.0	92%	.88	100%	1.0	90%	.84
10	100%	.*	100%	.*	92%	.81	92%	.79	100%	1.0	90%	.73
11	91%	.81	91%	.82	75%	.63	75%	.59	70%	.55	70%	.56
12	91%	.62	91%	.74	83%	.71	83%	.73	70%	.52	90%	.84
13	100%	1.0	100%	1.0	75%	.47	92%	.84	90%	.81	80%	.61
14	100%	.*	100%	.*	92%	.*	92%	.*	100%	1.0	100	1.0
15	100%	1.0	100%	1.0	83%	.75	100%	1.0	90%	.83	90%	.83
16	100%	1.0	82%	.66	75%	.60	92%	.84	90%	.82	70%	.47
17	91%	.84	100%	1.0	100%	1.0	92%	.88	90%	.81	60%	.48
18	100%	1.0	91%	.85	100%	1.0	100%	1.0	100%	1.0	100%	1.0
19	100%	1.0	100%	1.0	92%	.85	100%	1.0	100%	1.0	100%	.75
20	100%	1.0	91%	.83	92%	.87	92%	.87	90%	.83	90%	.83

Table 6. Weighted Kappa values and percentages of agreement for Non-Meaningful Gesture Test, first and second reliability study in typically developing (TD) children and in children with DCD.

*= not able to calculate since the item is an unvaried variable