Hill, Elisabeth L., Bishop, Dorothy V. and Nimmo-Smith, Ian

Representational gestures in Developmental Coordination Disorder and specific language impairment: Error-types and the reliability of ratings

ELSEVIER

Human Movement Science 17 (1998) 655–678

HUMAN MOVEMENT SCIENCE

Representational gestures in Developmental Coordination Disorder and specific language impairment: Error-types and the reliability of ratings

E.L. Hill [a,*], D.V.M. Bishop [b], I. Nimmo-Smith [b]

[a] *Department of Experimental Psychology, Downing Street, Cambridge CB2 3EB, UK*
[b] *MRC Applied Psychology Unit, 15, Chaucer Road, Cambridge CB2 2EF, UK*

Abstract

Gesture production was studied in children with developmental coordination disorder (DCD) and specific language impairment (SLI) in comparison to normally developing age-matched and younger control children. The pattern of error production was investigated to characterize the praxis skills seen in these two developmental disorders as well as to inform understanding of the aetiology of both DCD and SLI. Given the subjective nature of the categorization of errors, a separate study was conducted to investigate inter-rater reliability. Independent adult raters consistently used four out of fourteen error-types and for these four, inter-rater reliability was found to be good. The type of errors made by children with DCD, SLI and the younger controls were very similar. The only difference between the groups was in the frequency with which errors were displayed, suggesting that the performance of the clinical children might be an indicator of immature praxis development. It is suggested that the inclusion of a younger control (similar to a "motor match") group is an important methodological device for investigating the underlying nature of disorders such as DCD and SLI.

PsycINFO classification: 3250; 3270

Keywords: Coordination; Dyspraxia; Language disorders

* Corresponding author. Fax: +44 1223 333564; e-mail: eh215@cus.cam.ac.uk.

0167-9457/98/$19.00
PII: S 0 1 6 7 - 9 4 5 7 (9 8) 0 0 0 1 7 - 7

1. Introduction

Conventional medical classification systems subdivide specific developmental disorders into distinct categories, of which developmental coordination disorder (DCD) is one subtype. DCD is diagnosed in children who experience movement difficulties out of proportion with their general development, and in the absence of any known medical condition (e.g., cerebral palsy) or identifiable neurological disease (American Psychiatric Association, 1994). However, motor problems have also been described in other developmental disorders. Children who are not diagnosed on the basis of motor impairments have been shown to have concomitant difficulties on certain tasks involving motor control (e.g., children with verbal sequencing deficits – Dewey et al., 1988; developmental dyslexia – Nicolson and Fawcett, 1994; attention deficit hyperactivity disorder – Whitmont and Clark, 1996).

There are two possible accounts of this systematic overlap. The first maintains that the sharp distinction between developmental disorders that is made in textbooks may be artificial. Rather than there being discrete groups of children, some with motor coordination problems, some with reading problems, others with language or attentional problems, it may be that all these impairments tend to co-occur in developmentally disordered children, and that those with highly specific deficits are the exception rather than the rule. An alternative view is that there are clear cut distinctions between developmental disorders, and that the similarities in motor impairments are only superficial. According to this view, more detailed analysis should reveal qualitative differences in motor skills between children who meet criteria for DCD and children with other types of disorder.

In the current study, we undertook just such an analysis by comparing qualitative aspects of motor performance in children with DCD and with specific language impairment (SLI). In the same way as children with DCD fail to develop motor skills at a normal rate, children with SLI fail to develop language at a normal rate, for no apparent reason. The language functioning of such children is significantly below age level and out of proportion with the rest of their development. This impairment cannot be explained by physical impairments (e.g., abnormal articulators or hearing loss) or by identifiable neurological disease. Children with SLI were chosen as a comparison to those with DCD because they have been shown to have significant difficulty with certain types of motor and visuo-perceptual tasks (Archer and Witelson, 1988; Johnston et al., 1981; Powell and Bishop, 1992), reminiscent of the difficulties seen in children with DCD. Despite the fact that children with SLI have been shown to have movement difficulties resembling those seen in DCD, no direct comparison has been made between these two disorders.

We compared children with DCD and SLI on a test involving the production of representational gestures. The term "representational gesture" is applied to a number of tasks in which the participant demonstrates familiar actions. These can be either transitive (requiring the use of an object, e.g., comb your hair with a comb) or intransitive (movements that do not require an object, e.g., salute). Actions can be elicited in different response conditions, to verbal command (miming the action in the absence of the object) and imitation (copying the experimenter miming the action), for example. For adults with acquired apraxia, a typical performance profile sees transitive gestures performed more poorly than intransitive gestures, and all gestures performed more poorly to verbal command than to imitation.

Only a limited number of studies have investigated the ability to produce representational gestures in children. Perhaps not surprisingly, these have shown increased performance accuracy with age. The quantitative pattern of performance seen in adults, with transitive gestures performed more poorly than intransitive gestures and gestures to verbal command more poorly than to imitation, is observed in healthy children (Kaplan, 1968; Kools and Tweedie, 1975; Overton and Jackson, 1973), as well as in children with various disorders such as developmental motor deficits (Dewey, 1993; Dewey and Kaplan, 1992), sensorimotor dysfunction (Dewey, 1991) and learning disabilities (Cermak et al., 1980). Children with developmental disorders perform all gestures and all conditions significantly more poorly than their normally developing peers.

The results from a quantitative analysis of data from the present study have been reported elsewhere (Hill, 1998). This confirmed that children with DCD and those with SLI performed more poorly than age-matched controls on a representational gestures task, when actions were scored simply as correct or incorrect. This analysis showed that when performing to verbal command, children with DCD, SLI and the younger controls produced significantly more errors than children in an age-matched control group, both for transitive and intransitive actions. For gestures to imitation, however, the results were different for transitive and intransitive categories. For transitive gestures to imitation, children with DCD, SLI and younger controls performed less well than their normally developing peers. The only other result of significance for the transitive gestures was that the children with DCD produced more errors than those in the younger control group. For intransitive gestures to imitation, only children with DCD produced significantly more errors than their normally developing peers.

In the current study, our concern is with the qualitative characteristics of apraxic responses, rather than merely the speed or accuracy of performance. This has been much studied in relation to acquired apraxia in adults, for whom a rich taxonomy of error-types has been developed (e.g., Duffy and Duffy, 1989; McDonald et al., 1994). Indeed, the qualitative nature of performance has been argued to differentiate, in some cases, between different forms of apraxia in adults (e.g., Heilman, 1973; see Mozaz, 1992 for a review). The qualitative nature of gesture production of children, however, has not been addressed in such detail, and it is unclear whether the error-types observed in apraxic adults are present in children with developmental disorders or not. This question is of particular interest in relation to DCD, because this disorder is regarded by some as a form of developmental dyspraxia (e.g., Cermak, 1985; Dewey, 1995), i.e., a difficulty in programming movements that is not due to more fundamental limitations of sensory processing or motor control. If this conceptualization of DCD is accurate, we would expect to see classic apraxic errors on a representational gestures task. Moreover, since the motor difficulties of children with SLI have been described as arising principally from a slow rate of processing (Johnston et al., 1981), attentional limitations or lack of persistence (Powell and Bishop, 1992), we might expect the performance of children with DCD to be qualitatively different from that of children with SLI.

The first aim of this study, therefore, was to determine whether children's errors on a representational gesture task resembled those produced by adults and could be classified in terms of the same kind of framework. Since this kind of classification has not previously been undertaken, it was essential to check the reliability of such judgements. To achieve this we asked six adult raters who were naïve to the purposes of the experiment to rate the performance of a selection of children from our DCD, SLI and control groups.

The second aim of the study was to compare the qualitative nature of performance of children with SLI and those with DCD on the task of representational gestures. Insofar as errors by children with SLI do resemble those of children with DCD, we may also ask whether this is confined to those children who have other evidence of impaired motor performance. Since motor status was not a selection criterion for children in this group, we were also able to address the initial question of how many children with SLI would show impaired movement ability.

Of particular interest in the current study was the question of whether less accurate performance on a representational gestures task could be explained in terms of immaturity of a developing system or was more likely to be "pathological" in origin. To study this we included a younger control group of normally developing 5–6 year old children, matched to the older children on the basis of non-verbal IQ, age-appropriate language and to the clinical groups on the additional measure of peg-moving speed. This group acted as a form of "motor-age" match to the children in the clinical groups, analogous to the inclusion of a reading-age matched control group as used in studies of reading disability (cf. Goswami and Bryant, 1990) or more rarely of children with DCD (Hulme et al., 1984). The strength of this design lies in the opportunity it affords to determine whether the performance of impaired children resembles that of younger but normally developing children, who are less well developed motorically than the children in the age-matched control group, or whether it is qualitatively different. If similar performance was observed it would suggest that children with DCD and SLI have a maturational delay in the development of their praxis.

In summary, the main purposes of the present study were: (i) to ascertain whether it was possible to establish a reliable method of classifying qualitative aspects of children's representational errors, (ii) to compare the praxis skills of children with DCD and SLI using a qualitative error analysis and to consider whether either group made errors reminiscent of those made by adults with acquired apraxia, and (iii) to compare the two clinical groups with a younger control group with less mature motor skills, to consider whether the praxis errors seen in the clinical groups could be considered as developmental immaturities.

2. Method

2.1. Participants

A total of 72 children participated in this study. Three groups ranged in age from 7 to 13 years: (i) children with developmental coordination disorder (DCD), (ii) children with specific language impairment (SLI), and (iii) age-matched control children. A fourth group of younger control children (YC) ranged in age from 5 to 6 years. Performance on a number of assessments were used as inclusion/exclusion criteria (see below for detail).

2.1.1. DCD group

Eleven children referred to Child Development Centres in East Anglia and the south-east of England because of movement difficulties were recruited for this study. Eight were boys and three were girls (mean age = 9 years 3 months). To be included in this group the following inclusion/exclusion criteria were applied: a Movement ABC (Henderson and Sugden, 1992) score at or below the 15th percentile, a standardized non-verbal IQ score above 80 on Raven's Progressive Matrices (Raven et al., 1986) and a standardized score above 80 on the CELF-R Repeating Sentences (Semel et al., 1980). No child showed evidence of neurological impairment.

2.1.2. SLI group

Nineteen children with SLI (12 male; 7 female) were recruited from residential schools for children with SLI in East Anglia and the south-east of England. All pupils at these schools have comprehensive psychological and medical evaluations prior to school entry, and only those with severe and selective language difficulties are enrolled. Children with neurological impairment, physical handicap, hearing loss and those for whom English was not the first language spoken at home were excluded from the study. To be included in the sample, children had to achieve a non-verbal IQ above 80 on Raven's Progressive Matrices and a standardized score of 80 or below on the CELF-R Repeating Sentences. In this group, the mean age was 9 years 9 months.

Following initial selection, this group was then subdivided into an SLI-Clumsy and SLI-Pure group, according to a child's performance on the Movement ABC. Using the same criterion as for DCD, 11 children with SLI obtained scores at or below the 15th percentile on the Movement ABC (mean ABC score = 19.0), these children were termed "SLI-Clumsy". The remaining eight children with pure SLI (termed the "SLI-Pure" group) performed above the 15th percentile on the Movement ABC (mean ABC score = 5.1).

2.1.3. Age-matched controls

A group of twenty five control children (14 male; 11 female) were selected individually from primary schools in Cambridge to match the DCD and SLI groups as closely as possible on age, sex and non-verbal IQ. To be included in the group, children had to achieve a non-verbal IQ score above 80 on Raven's Progressive Matrices, a standardized score above 80 on the CELF-R Repeating Sentences and a score above the 15th percentile on the

Movement ABC. In fact, while performance of this group on the Movement ABC ranged from the 16th to the 96th percentile, the mean score fell above the 70th percentile (see Table 1). This group ranged in age from 7 to 12 years (mean = 9 years 8 months).

2.1.4. Younger controls

Seventeen younger control children (9 male; 8 female), aged 5–6 years (mean = 5 years 8 months) were selected individually from a local primary school and were matched by sex and non-verbal ability to children in the DCD, SLI and age-matched control groups. Since the CELF-R and Raven's Progressive Matrices are not suitable for children under 7 years, the non-verbal ability and language levels of children in this group were assessed using subtests from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1990). All children scored above the 15th percentile on the Movement ABC. While the performance of this group ranged from the 18th to the 93rd percentile, the mean score fell above the 62nd percentile.

Although the motor test employed in this study indicates clearly how a child compares to his/her own age peers, it is difficult to compare children across a wide range of ages as the test items change as children get older. In order to be more confident that our impaired children were well matched in their level of motor performance to our younger control children, we de-

Table 1
Group means (standard deviation) for age, the scaled scores of non-verbal and language ability and for the total scaled scores of the Movement ABC

	DCD ($n = 11$)	SLI ($n = 19$)	Control ($n = 25$)	YC ($n = 17$)	F ratio
Age (year)	9.3_a (1.4)	9.9_a (1.9)	9.8_a (1.5)	5.8_b (0.4)	25.11, $p < 0.001$
Non-verbal ability [a]	-0.2_a(0.8)	0.1_a (1.0)	0.5_a(0.7)	0.4_a(1.1)	1.41, $p > 0.1$
Language [b]	-0.1_a (1.0)	-2.7_b (0.4)	0.3_a (0.6)	0.2_a (1.1)	66.37, $p < 0.001$
Movement ABC [c]	19.6_a(6.1)	13.1_b(8.9)	2.1_c (2.9)	2.4_c (2.9)	36.64, $p < 0.001$
ABC Range	11.5–29.5	2–30.5	0–9.5	0–9.5	

Note: Means with different subscripts differ significantly at $p < 0.05$ by the Fisher Least Significant Difference test.

[a] Group matching test: scaled scores are given. All scored within the normal range.

[b] Group selection test: scaled scores are given. SLI scored below the normal range; other groups scored within the normal range.

[c] Group selection test: Total scaled scores are given. A high score indicates impairment. Different tests are completed according to a child's age. DCD scored at or below 15th percentile (raw score of 10 or above); control scored above 15th percentile; SLI not selected for this: 11 out of 19 children scored at or below 15th percentile.

cided to use one test which could be administered to all children. It is difficult to find a motor test that is suitable for this purpose but after pilot work we chose a task that involved fast, precise movements of the hands that is known to be easy for most 7–8 year old children. This was the "Placing Pegs" item from age band two of the Movement ABC in which the child must place twelve plastic pegs into a peg board as quickly as they can, with each hand separately. The procedure was identical to that described on page 58 of the Movement ABC manual. Both hands were tested and a practice trial was given for each hand. For the purposes of the current study, the number of seconds taken to complete the task was recorded and speed of peg-moving was collapsed for the two hands. Since the appropriate age band of the Movement ABC had been administered to all of the children as part of the selection procedure (see Section 2.2), these data were already available for all of the children aged 7–8 years in the clinical and age-matched control groups. All of the children in the younger control group were then given this peg-moving task in addition to the age-appropriate items of the Movement ABC used in the selection process.

Using this procedure, six children with DCD, eight with SLI, and seven age-matched controls could be compared with the 17 younger control children. There was no significant difference in completion time between the different groups (see Table 2) and we were thus confident that the groups were well matched in terms of their levels of motor competence.

The means and standard deviations for age, the scaled scores of non-verbal and language ability, along with the total scaled scores for the Movement ABC for all groups are shown in Table 1. There was no significant gender difference between any of the groups (F (3,68) = 0.43, $p > 0.1$).

Table 2
Group means (standard deviation) for speed (s) of completing the "motor match" (peg-moving) test of the Movement ABC (age band two)

	DCD ($n = 11$)	SLI ($n = 19$)	Control ($n = 25$)	YC ($n = 17$)	F ratio
Pegmoving time [a]	30.0_a (5.2)	25.7_a (4.9)	22.1_b (2.0)	29.3_a (5.4)	4.49, $p < 0.001$

Note: Means with different subscripts differ significantly at $p < 0.05$ by the Fisher Least Significant Difference test.

[a] Pegmoving time (s) for all younger controls on the pegmoving task for 7–8 year olds on the Movement ABC (Band 2) and for the children in the clinical and control groups aged 7–8 years (SLI $n = 8$; DCD $n = 6$; Control $n = 7$). Data is collapsed across the preferred and non-preferred hands.

2.2. Procedure

Once the study had been approved by the Ethics Committees of Cambridge and Worthing District Health Authorities, the various schools and clinics involved in the study were approached for potential participants. All children were then assessed in their clinics or schools. First, a number of experimental tests, including the gesture production test (other tests are reported elsewhere; e.g., Hill, 1997, Hill and Bishop, in press), were completed along with the tests of non-verbal and language abilities. Second, the Movement ABC test was completed.

2.2.1. Selection tests

Assessment of non-verbal intelligence. In view of the language difficulties experienced by the children in the SLI group, a non-verbal intelligence test was used to ensure that all children were given an equal opportunity to demonstrate their reasoning ability. Two different tests were used depending on the age of the children, each administered by EH. Raven's Progressive Matrices was used for children in the DCD, SLI and age-matched control groups. This is a test which does not require a verbal response and is uninfluenced by manual dexterity. Test-retest reliability is reported as 0.88. The 1982 standardization was used to derive norms on the Coloured Matrices for children aged 7 years 0 months to 11 years 9 months and the 1979 standardization for Progressive Matrices for children aged 11 years 9 months and older. Scores were converted to scaled scores using a mean of 100 and SD of 15. The Picture Completion subtest from the WPPSI was used to assess non-verbal ability in the younger children. This is a task in which children must identify what is missing from a series of pictures. Test-retest reliability is reported as 0.82. Scores were converted to scaled scores using a mean of 10 and SD of 3.

Assessment of language. Two different language tests were used depending on the age of the children, each administered by EH. The Repeating Sentences subtest from the Clinical Evaluation of Language Fundamentals-Revised (CELF-R) was used with the DCD, SLI and age matched controls. It was selected because Bishop et al. (1995) found it to be particularly sensitive to SLI. Repeating Sentences tests auditory-verbal memory for sentences of increasing grammatical complexity. The experimenter reads out a sentence, which the child must then repeat. CELF-R has not been standardized officially in the UK, although Bishop et al. (1995) reported that British children scored similarly to the US standardization sample. No data on reliability are reported in the test manual. Scores were converted to scaled scores using a mean of 100 and SD of 15. For the younger controls, the Repeating Sentences subtest from the WPPSI was used to assess the same elements of language ability as the CELF-R. This subtest is administered in the same way as CELF-R and test-retest reliability is reported as 0.79. Scores were converted to scaled scores using a mean of 10 and SD of 3.

Assessment of motor competence. Motor competence of children in all groups was assessed using the Movement ABC administered by EH. This is a test battery designed to identify children with impaired motor development. A total of eight tasks measuring manual dexterity, ball skills and balance are completed. Subtests are slightly different according to a child's age, but address the same aspects of motor skill. Each test is scored on a scale of 0–5, a higher score indicating a less motorically competent child. The Movement ABC has been standardized in the US. Overall reliability is considered to be good, ranging from 97% in 5 year old children to 73% in 9 year olds. In the current study the age-appropriate band of the Movement ABC was used for each child. Scores on the individual items of the test are summed to produce a total ranging from 0 to 40. These are then converted to percentiles according to norms presented in the manual.

2.2.2. Test of gesture production

As noted above, there is a substantial literature concerning gesture production. We used the gestures described in Dewey (1993). Children sat at a table opposite the experimenter. They were asked to demonstrate six transitive gestures (brush teeth with a toothbrush, comb hair with a comb, eat ice cream with a spoon, hit a nail with a hammer, cut paper with scissors, write with a pencil) and six intransitive gestures (salute, wave good-bye, blow a kiss, make a fist, snap fingers, show that you have a full stomach). Each gesture was completed to verbal command (VC) and imitation (IM). The order of conditions (transitive-VC, transitive-IM, intransitive-VC, intransitive-IM) was randomized and counterbalanced across participants. Performance was videotaped for all children.

A list of possible errors was compiled from previously reported studies of the gesture production of adult neuropsychological patients (see McDonald et al., 1994 and Mozaz, 1992 for reviews). These are listed and described briefly in Table 3. The gestures of each child were observed on video by the first author (EH) who completed a checklist of errors for each action in each condition.

Table 3
Description of error-types

Error name	Description (where necessary)
Amorphous	Movement bears no resemblance to that requested
Body-part-as-object	Part of the anatomy is used to represent the object, e.g., index finger used to represent toothbrush
Clumsy	Clumsy or awkward movement
Delayed initiation	Correct movement is shown after a delay
External configuration	Correct grip of an object is shown, but the length of the object is not accounted for in the action, e.g., placing the hand too close to the mouth when miming using a toothbrush
Internal configuration	Correct use of an object is shown, but the object is oriented incorrectly, e.g., when demonstrating the use of a comb, the hand grip would not allow the comb to come into contact with the hair
"I don't know"	
No response	
Perseveration	Child repeats previous action
Spatial orientation	Hand deviates from the appropriate spatial position
Substitution	Discrete part of action bears no resemblance to that requested
Target mislocation	Gesture not performed on the correct body part
Unsustained	Movement starts accurately, quickly deteriorates
Vocalization	Verbal description accompanies movement

2.2.2.1. Inter-rater reliability study. In order to check the reliability of the error analysis, an edited video was compiled containing 84 representative examples of the gestures produced by 20 children in the study. The tape included examples from children in each of the four groups as well as examples of performance under each of the experimental conditions.

Six adult raters (three male and three female) aged between 27 and 55 were recruited from the volunteer subject panel of the Applied Psychology Unit, Cambridge, UK. All were blind to the purposes of the study, and none had previous experience of the literature relating to gesture production. Before watching the video, the raters were told that the aim of the study was to determine how people classify the movements that children make when they are asked to mime certain actions. The 12 actions (transitive and intransitive) were then demonstrated to them by the experimenter. Precise details of the different error-types that they might see were then provided. These included the 14 error-types used by the experimenter to assess the data (see Table 3) as well as a "correct performance" response.

After watching an action on the video, the raters were asked to write down which of the errors were appropriate for the gesture that they had seen. Several points were stressed at this stage: (i) more than one error could be appropriate for each action, (ii) not every error had to be used, and (iii) correct performance was a possible response. Before each gesture was viewed on the video, raters were told which action would be seen, as well as its response condition (verbal command/imitation). Ample time was given for responding and, if necessary, actions were reviewed on the video.

3. Results

3.1. Qualitative analysis of the children's errors

The first objective of this study was to examine the nature of the errors made by the children and compare them to those known to be produced by adults. The outcome of this analysis is summarised in Table 4. Some errors were not mutually exclusive and several actions were assigned more than one error. For example, an action could be described as both ‘‘unsustained'' and showing an ‘‘external configuration'' response.

Table 4
Proportion of children showing at least one example of each error-type, as rated by the experimenter

	DCD ($n=11$)	SLI ($n=19$)	Control ($n=25$)	YC ($n=17$)
Amorphous movement	0.5	0.1	0.1	0
Body-part-as-object	0.8	0.9	0.6	1
Clumsy movement	0.3	0.2	0	0.2
Delayed initiation	0	0	0	0
External configuration	0.8	0.8	0.6	0.9
Internal configuration	0.7	0.5	0.2	0.4
‘‘I don't know''	0.2	0.3	0	0.3
No response	0.2	0	0	0
Perseveration	0	0	0	0
Spatial Orientation	0.7	0.6	0.4	0.7
Substitution	0	0	0	0
Target mislocation	0.1	0.2	0	0
Unsustained action	0	0.1	0	0
Vocalization	0.1	0.1	0	0

Note: Fourteen possible error-types are shown. In addition, a classification of ‘‘correct performance'' could be made.

3.2. Inter-rater reliability

Moving back to the inter-rater reliability study, two versions of the Kappa statistic were used to establish whether there was consistency of response for each action across the raters. The two-rater Kappa Coefficient (Cohen, 1960; Fleiss, 1981) and the Multiple-Rater Kappa (Fleiss, 1981) are measures of the overall agreement between a group of two or more raters. Both of these measures adjust the amount of observed agreement for the amount that might have occurred by chance if raters were assigning errors independently. As well as assessing the reliability of the entire set of errors by means of these coefficients, it is possible to calculate how much chance-corrected agreement there is between two or more raters in their use of each single category in turn. Across the adults who rated the children's data, 26 combinations of error were noted, which when added to the number of "pure" errors provides a total of 41.

When all 41 combinations of errors were considered, there was a significant amount of agreement between the six raters (Multiple Kappa = 0.3 (SE = 0.03), $Z = 9.64$, $p < 0.001$). This is not a strong level of agreement, as there was not a particularly strong consensus as to which errors applied. Of the 84 actions viewed on video, all six raters agreed on which error applied to 27 actions (32%), a further five actions (6%) gained consensus from five raters, and 12 actions (14%) received a common error rating from four raters.

However, an examination of the original 14 individual error-types revealed that only seven had multiple Kappas above 0.1 (see Table 5), as did the category of "correct". For the purposes of the current paper, analysis focused on the ratings of five key errors ("correct", body-part-as-object, external configuration, internal configuration, spatial orientation). "Correct", "body-part-as-object" and "external configuration" were selected because they showed good inter-rater reliability (Kappa values of 0.58, 0.58 and 0.42, respectively). Although Kappa values for "internal configuration" and "spatial orientation" were not high, these errors were chosen for analysis because they were common across all groups of children as well as subjectively appearing prevalent during testing sessions. While the error-types "I don't know" and "vocalization" have good Kappa values (0.64 and 0.52, respectively), these errors were not included in the analysis because inter-rater agreement was inevitable by virtue of their definition (if children do no know what an action is then they cannot respond and therefore good agreement will be seen among raters). Mislocation also had a Kappa value above 0.1, but was not considered because this error-type was produced only by one child.

Table 5
Multiple kappa (standard error) values for the original 14 error-types, as well as for "correct performance", derived from the responses of the six naïve adult raters to the video of the actions produced by the children

Error name	Kappa (SE)
Correct	0.58 (0.03)
Amorphous	0.01 (0.49)
Body-part-as-object	0.58 (0.1)
Clumsy	0.04 (0.09)
Delayed initiation	−0.01 (0.4)
External configuration	0.42 (0.1)
Internal configuration	0.14 (0.16)
"I don't know"	0.64 (0.27)
No response	−0.01 (0.57)
Perseveration	–
Spatial orientation	0.11 (0.12)
Substitution	–
Target mislocation	0.13 (0.57)
Unsustained	0.04 (0.27)
Vocalization	0.52 (0.2)

Note: Perseveration and substitution were not seen in the children's responses.

A revised multiple Kappa analysis was conducted by refining the dataset to consider inter-rater reliability only for the four error-types which would be considered in the analyses reported below. The "correct" response was also included in this refined multiple Kappa analysis. In addition, a category of "other" was included, comprising all observations made which did not include any of the five key errors as part of the error combinations. Following the repeated Multiple Kappa analysis, the responses of the raters showed an acceptable amount of agreement (Multiple Kappa = 0.41 (SE = 0.03), $Z = 13.96$, $p < 0.001$). The consensus is exhibited by all six raters assigning 30 actions (36%) to a single error-type, and a further 30 actions (36%) getting majority assignments from 4 or 5 raters.

In sum, the data emerging from these analyses has demonstrated that children's errors on the gesture production task can be classified using a scheme derived from the adult literature. Although some of the error-types initially explored were not used very frequently, others were used often enough to be useful and could indeed be rated reliably. For the main part of the study, reported below, we focus attention on four categories of error, body-part-as object (BPO), external configuration (EC), internal configuration (IC) and spatial orientation (SO).

3.3. Qualitative differences between the groups of children

Some of the categories identified in the adult apraxic literature were not applicable to the responses given by any of the children. The proportion of children showing at least one example of each error-type is shown in Table 4. It can be seen that errors of "perseveration" and "substitution" were not noted for any child.

3.3.1. Group comparisons

Group comparisons were made between five groups of children: the DCD group, the children with SLI spilt into those who were "clumsy" and those who were not, age-matched and younger controls. The frequency of the four error-types under consideration is shown in Fig. 1 as a percentage of the total

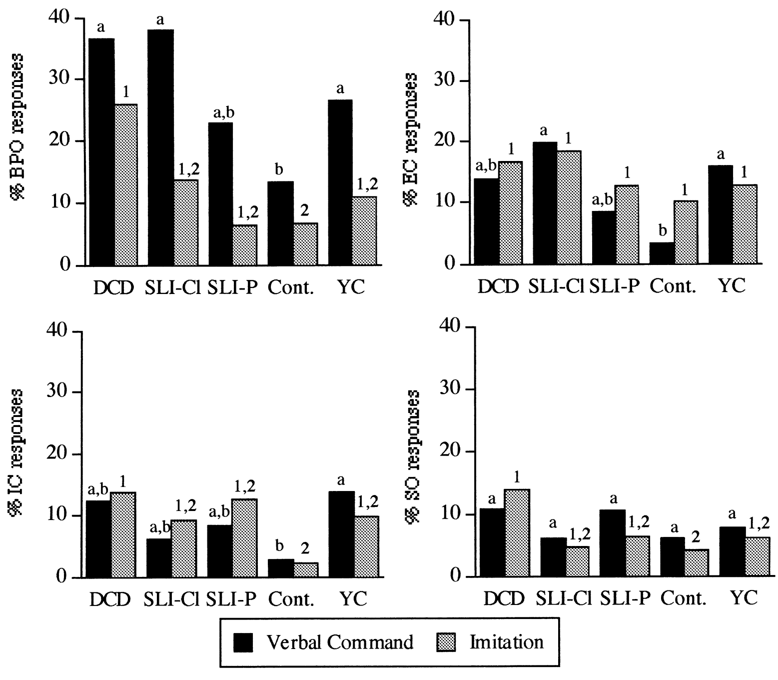

Fig. 1. Percentage of responses of each error-type (BPO, EC, IC, SO) for each group, with the SLI group split into SLI-Clumsy (SLI-Cl) and SLI-Pure (SLI-P). Data for BPO, EC and IC errors are from performance of transitive gestures only. Data for SO errors are from performance of intransitive gestures only. For verbal command, columns with different letters differ significantly, while columns with different numerals indicate differences when imitating.

number of responses of each group. Mann Whitney *U* tests were performed to compare the number of errors produced by the children in each group. Owing to the multiple comparisons being made, probabilities were derived using Bonferroni adjustments, and a significance level of $p < 0.01$ was taken. The nature of the group comparisons are depicted in Fig. 1, showing that, in all cases, there was at least a trend towards the children with DCD, SLI-Clumsy, SLI-Pure and the younger controls producing more of each of these errors than the age-matched controls. Of particular note is the non-significant trend for children with pure SLI to make more errors than age-matched controls for all error-types, as well as the fact that no comparisons between those in the SLI-Clumsy and SLI-Pure groups reached statistical significance.

The results of the planned comparisons produced three important results. First, children in all groups produced errors of similar types and there were differences only in the frequency with which they were observed. In all cases, gestures were correct conceptually, i.e., they resembled the correct action. Second, children in the DCD, SLI and younger control groups performed similarly to one another but differently from the control group matched for age. Taking a level of significance of $p < 0.01$ meant that although group comparisons were not always significant at the chosen alpha level, the trend was there in other cases (see Fig. 1). Finally, performance to imitation reduced, but did not eliminate, error production.

It could be argued that poor performance to verbal command arose from incomprehension of the instructions by certain children. This could have been a problem for the children with SLI in particular. If this were the case, one would predict that such children would perform well to imitation since a correct copy of the experimenter's action could be completed without comprehension of verbal instruction. Fig. 1 shows that errors were made by all children in both response conditions. In addition, the fact that errors were correct conceptually to both verbal command and imitation suggests that poor performance cannot be explained on the basis of poor comprehension.

3.3.2. Action specific errors

In developing an error classification, it quickly becomes apparent that different actions allow for different types of error. Of the predominating error-types, BPO, EC and IC ratings could be assigned only to transitive actions. Whereas SO ratings tended to be assigned only to the intransitive gestures. Furthermore, within the transitive actions, BPO responses tended to be used for the scissors and pencil, whereas EC responses predominated when miming toothbrushing.

This raises the question of how far the error patterns shown in Fig. 1 reflect the influence of specific problematic items. Given that there were no significant differences between the performance of the SLI-Clumsy and SLI-Pure subgroups, their data were collapsed, and an investigation of item specificity was conducted on the data for the four original groups. Collapsing the SLI subgroups had the advantage of avoiding the problem of the small sizes of these subgroups. The percentage of each group who responded using the four main errors for each action and response condition (verbal command/imitation) is shown for transitive and intransitive actions in Tables 6 and 7, respectively.

In scanning Tables 6 and 7, it is clear that in most instances where the DCD group made a high proportion of errors, so too did the younger control children, e.g., BPO when miming the use of a toothbrush. However, there are some intriguing exceptions, and these tended to occur in the imitation conditions. It is difficult to uncover a common pattern across actions, but inspection of predominant error-types for a given action can be informative. For instance, on the "spoon" item, we can see that the DCD, SLI and younger control groups make a relatively high proportion of EC errors to verbal command. On imitation, this pattern persists for the SLI and younger control

Table 6
Percentage of each group making BPO, EC and IC errors, separately for each transitive action in each response condition (verbal command, VC/imitation, IM) (each action is denoted by the object whose use was mimed)

	% BPO				% EC				% IC			
	DCD	SLI	CON	YC	DCD	SLI	CON	YC	DCD	SLI	CON	YC
TRANS-VC												
Toothbrush	45.5	36.8	4	11.8	54.6	26.3	12	35.3	0	5.3	0	0
Comb	18.2	15.8	12	5.9	0	0	0	0	36.4	10.5	0	23.5
Spoon	0	0	0	0	63.6	63.2	8	52.9	18.2	0	0	0
Hammer	45.5	36.8	4	52.9	0	0	0	0	9.1	21.1	16.0	0
Scissors	63.6	63.2	40	64.7	0	0	0	0	0	0	0	0
Pencil	27.3	36.8	20	23.5	0	0	0	0	0	0	0	0
TRANS-IM												
Toothbrush	36.4	15.8	4	0	36.4	42.1	56	23.5	0	10.5	0	0
Comb	9.1	0	8	5.9	9.1	0	4	0	27.3	15.8	0	11.8
Spoon	0	0	0	0	18.2	52.6	4	52.9	45.5	5.3	0	5.9
Hammer	45.5	10.5	0	29.4	0	0	0	0	0	26.3	12	5.9
Scissors	45.5	15.8	16	23.5	0	0	0	0	0	0	4	0
Pencil	36.4	21.1	12	0	0	0	0	0	0	0	0	0

Table 7
Percentage of each group making SO responses, separately for each intransitive action in each response condition (verbal command, VC/imitation, IM)

	% SO							
	DCD	SLI	CON	YC	DCD	SLI	CON	YC
INTRANS-VC					*INTRANS-IM*			
Salute	63.6	42.1	36	47.1	81.8	31.6	24	29.4
Wave	0	0	0	0	0	0	0	0
Blow kiss	0	5.3	0	0	9.1	5.3	0	0
Fist	0	0	0	0	0	0	0	0
Click fingers	0	0	0	0	0	0	0	0
Full stomach	0	0	0	0	0	0	0	0

groups, but the predominant error for DCD shifts to IC. These observations are tentative, of course, but they do suggest that imitation tasks might be more sensitive in eliciting errors which go beyond developmental immaturities.

This descriptive account of the patterns of errors can be reinforced by statistical analyses. Though there is not yet available the equivalent of a repeated measures analysis of variance for multiple category data, it is possible to assess the differences between the various actions in the prevalence of each type of error, and whether the profile across the actions varies with the subject group. To this end, three separate linear logistic regression analyses (Collett, 1991) were applied to the data for transitive actions produced to verbal command (for the presence or absence of each of the three types of error BPO, EC and IC; see Table 6). The results for the imitation data are similar and are not reported here. It should be noted that for intransitive actions errors were seen almost exclusively on the action "salute" (see Table 7) and consequently these data were not analysed using the linear logistic regression method.

For BPO, the effect of group was significant (deviance $\chi^2 = 17.2$, df = 3, $p < 0.001$), with the control group producing significantly fewer BPO responses. There was a significant main effect of action (deviance $\chi^2 = 111.2$, df = 5, $p < 0.001$), which is accounted for by the higher rate of BPO errors when miming the use of "scissors". There was no evidence of a group × action interaction (deviance $\chi^2 = 20.8$, df = 15, $p > 0.1$), though the performance of children in the age-matched control group when miming the use of a pencil might be considered to be unexpectedly poor.

For EC, the effect of group was significant (deviance $\chi^2 = 19.1$, df = 3, $p < 0.001$), with the control group producing significantly fewer EC responses. There was a significant main effect of action (deviance $\chi^2 = 168.0$, df = 5, $p < 0.001$), which is accounted for by the higher rate of EC errors when miming the use of a "spoon", followed by an intermediate rate for miming the use of a "toothbrush". There was no evidence of a group × action interaction (deviance $\chi^2 = 7.64$, df = 15, $p > 0.1$).

For IC, the effect of group was not significant (deviance $\chi^2 = 6.0$, df = 3, $p > 0.1$). The effect of action was significant (deviance $\chi^2 = 40.1$, df = 5, $p < 0.001$), with significantly more IC responses being produced when miming the use of a "comb" and "hammer" as opposed to the other four actions. There was no evidence of a group × action interaction (deviance $\chi^2 = 22.4$, df = 15, $p > 0.1$).

4. General discussion

The main purpose of the current study was to characterize the praxis skills of children with DCD and SLI in comparison to normally developing age-matched as well as younger (motor matched) control children and to use the pattern of error production on a gesture representation task to inform understanding of the aetiology of both DCD and SLI.

An essential element of the study concerned the issue of establishing a reliable method of classifying qualitative aspects of children's errors on the gesture representation task. An error categorization based on the adult apraxia literature proved a useful starting point. Some error-types, however, were never observed in the children's responses, and others could not be rated reliably. Children in each group produced errors of the same types to differing degrees. Children with DCD, SLI and the younger controls simply made more of these errors than children in the older control group. In all children, errors were correct conceptually, both when performing to verbal command and imitation, even when inaccuracies occurred. Where an error was produced, therefore, it seems reasonable to interpret these as originating from imprecise implementation of the sequence of movements, rather than an inability to conceptualize the action.

The finding that all children's errors were correct conceptually suggested that, rather than representing abnormal performance, the observed errors may arise from a continuum of performance, with the performance of the children in the clinical and younger control groups simply falling within the more error-prone range of normal performance. This suggestion is endorsed by studies showing a developmental progression in the ability of normally developing children to perform a representational gestures task. The effect of item-specificity, where specific actions elicited certain errors in all the groups, supports this view.

A central component of the current study concerned the nature of the error responses produced by each group of children. Kaplan (1968) and others (e.g., Kools and Tweedie, 1975; Overton and Jackson, 1973) have reported age-related changes in the nature of the responses produced on a representational gestures task. BPO was common in 4 year olds, but decreased with the emergence of symbolic representation of the imagined objects. Haaland and Flaherty (1984) reported that by the age of 8 years BPO use had declined, making way for a "less primitive" error-response, that of external configuration. In this study, the performance of the children in the younger control group supports such a picture, typically more errors were produced by this group than by the older controls. However, errors did not drop out totally in the latter group. Indeed, healthy adult controls have been shown to make BPO errors as often as patients with unilateral damage to the left and right hemispheres (Duffy and Duffy, 1989), which might suggest that the BPO errors produced by the older control children in the current sample reflect their own concept of symbolization. It is possible to imagine that use of BPO may arise because a person is wanting to perform an action that is easier for the viewer to interpret.

The current study adds to the debate about whether children with SLI show performance comparable to that of children with DCD. The performance of these two clinical groups was similar, a finding that has been suggested, but not investigated directly in the previous literature (e.g., Archer and Witelson, 1988; Johnston et al., 1981; Powell and Bishop, 1992). The question arises as to the nature of the SLI-motor association (which is perhaps similar to the frequent co-occurrence of acquired apraxia and aphasia). There are three possible explanations. First, the language disorder plays a specific role in the deficits seen in a representational gestures task. If this is the case, we would expect a high correlation between language impairment and motor performance, something which was not seen in the current study. Second, these deficits may be the consequence of the anatomical contiguity of the neural substrates subserving language and motor functions. Alternatively, the apparent relationship between language and motor difficulties may arise because both deficits are indicators of underlying immaturity of brain development (i.e., compromised nervous systems). If this is the case, it is likely that children with other developmental disorders (e.g., dyslexia, attention deficit hyperactivity disorder) will have similar difficulties on this test of gestural production. The limited evidence that is available on this issue suggests that this might indeed be the case (e.g., Dewey et al., 1988), which, in conjunction with the findings of the current study, is suggestive of overlapping, rather than distinct disorders, as well as being suggestive of a common aetiology.

Do we have evidence that performance was suggestive of "pathology" or of developmental delay in the clinical groups? There are two sources of evidence which can hint at this. First, in the current sample, children with SLI performed similarly to those in the younger control group which is consistent with an interpretation in terms of motor immaturity rather than brain damage. Certainly neurological immaturity has been suggested in the past as an explanation for the poor performance of children with SLI (Johnston et al., 1981; Bishop and Edmundson, 1987). Further support for such an interpretation comes from brain imaging studies which have provided some evidence of atypical morphological asymmetries in language-impaired children, but no evidence of a specific lesion site (e.g., Jernigan et al., 1991; Tallal and Katz, 1989). The cross-sectional nature of the current data do not allow us to propose the likelihood of the validity of one explanation over another. However, the inclusion of the younger control group in the current study adds impetus to the suggestion that immaturity rather than pathology may be the root of the performance observed in the gesture production of the children with SLI. Without this younger control group, which acted as a motor match to the children in the clinical groups, we would have been more likely to conclude that pathology was a plausible explanation for the error responses made by the children with SLI. For this reason, the current study provides a useful methodological strategy for future work.

In the children with DCD, the picture concerning the issue of pathology or developmental delay was less clear. In the verbal command condition, the performance of children with DCD was similar to that of the younger controls, but when imitating gestures, children with DCD were poorer. There is evidence that the difficulties experienced by children with DCD do not generally resolve by adolescence (e.g., Cantell et al., 1994; Losse et al., 1991). In addition, Jongmans et al. (1993) identified echodensities (flares) in the periventricular white matter of preterm infants. On follow-up after 6 years, those children whose flares had lasted more than 14 days exhibited significant motor delays compared to preterms without flares on earlier scans, or in those whose flares had lasted less than 14 days. In addition, other studies using

CT scans have shown that children with DCD typically show a higher incidence of cerebral abnormality than seen in their normally developing peers, but no specific lesion location (Knuckey et al., 1983).

In conclusion, this study presents a more extensive qualitative description and analysis of gesture production than reported previously in the developmental literature. We have shown that qualitative analysis is a useful tool if one wishes to obtain a more accurate picture of the nature of the errors made by children on a task of representational gestures and that this analysis can be undertaken reliably. We have also shown that the inclusion of a younger control group, which can act as a form of motor match to the children in the clinical groups, has the potential to inform work more than just a control group matched for chronological age. In the current study the performance of this younger control group in comparison to that seen in the clinical groups is suggestive of an immaturity in the development of praxis, rather than dysfunction. This is a preliminary study, however, and more detailed experimental research is needed, to examine this theoretical issue.

Acknowledgements

We would like to thank the children and their parents who participated in this project, as well as the occupational therapists and teachers who helped with recruitment. This study was completed in partial fulfillment of the requirements for a PhD. The work was supported by a Medical Research Council Studentship awarded to the first author and was conducted while she was based at the Medical Research Council's Applied Psychology Unit, Cambridge, UK.

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Author, Washington DC.

Archer, L.A., Witelson, S.F., 1988. Manual motor functions in developmental dysphasia. Journal of Clinical and Experimental Neuropsychology 10, 47.

Bishop, D.V.M., Edmundson, A., 1987. Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. Developmental Medicine and Child Neurology 29, 442–459.

Bishop, D.V.M., North, T., Donlan, C., 1995. Genetic basis of specific language impairment. Evidence from a twin study. Developmental Medicine and Child Neurology 37, 56–71.

Cantell, M.H., Smyth, M.M., Ahonen, T.P., 1994. Clumsiness in adolescence: Educational, motor and social outcomes of motor delay detected at 5 years. Adapted Physical Activity Quarterly 11, 115–129.

Cermak, S.A., 1985. Developmental dyspraxia. In: Roy, E.A. (Ed.), Neuropsychological Studies of Apraxia and Related Disorders. Elsevier, Amsterdam.

Cermak, S.A., Coster, W., Drake, C., 1980. Representational and nonrepresentational gestures in boys with learning disabilities. American Journal of Occupational Therapy 34, 19–26.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Collett, D., 1991. Modelling Binary Data. Chapman & Hall, London.

Dewey, D., 1991. Praxis and sequencing skills in children with sensorimotor dysfunction. Developmental Neuropsychology 7, 197–206.

Dewey, D., 1993. Error analysis of limb and orofacial praxis in children with developmental motor deficits. Brain and Cognition 23, 203–221.

Dewey, D., 1995. What is developmental dyspraxia? Brain and Cognition 29, 254–274.

Dewey, D., Kaplan, B.J., 1992. Analysis of praxis task demands in the assessment of children with developmental motor deficits. Developmental Neuropsychology 8, 367–379.

Dewey, D., Roy, E.A., Square-Storer, P.A., Hayden, D., 1988. Limb and oral praxic abilities of children with verbal sequencing deficits. Developmental Medicine and Child Neurology 30, 743–751.

Duffy, R.J., Duffy, J.R., 1989. An investigation of body part as object (BPO) responses in normal and brain-damaged adults. Brain and Cognition 10, 220–236.

Fleiss, J.L., 1981. Statistical Methods for Rates and Proportions. Wiley, New York.

Goswami, U., Bryant, P., 1990. Phonological Skills and Learning to Read. Lawrence Erlbaum, Hove.

Haaland, K.Y., Flaherty, D., 1984. The different types of limb apraxia errors made by patients with left vs right hemisphere damage. Brain and Cognition 3, 370–384.

Heilman, K.M., 1973. Ideational apraxia – a redefinition. Brain 96, 861–864.

Henderson, S.E., Sugden, D.A., 1992. Movement Assessment Battery for Children. The Psychological Corporation, Sidcup.

Hill, E.L., 1997. An Investigation of the Motor Deficits in Children with Developmental Co-ordination Disorder or Specific Language Impairment. Ph.D. Thesis, University of Cambridge, Cambridge.

Hill, E.L., 1998. A dyspraxic deficit in developmental coordination disorder and specific language impairment? Evidence from hand and arm movements. Developmental Medicine and Child Neurology, 40, 388–395.

Hill, E.L., Bishop, D.V.M. A reaching test reveals weak hand preference in specific language impairment and developmental coordination disorder. Laterality, in press.

Hulme, C., Smart, A., Moran, G., McKinlay, I., 1984. Visual, kinaesthetic and cross-modal judgements of length in normal and clumsy children: A comparison with young normal children. Child: Care Health and Development 10, 117–125.

Jernigan, T., Hesselink, J.R., Sowell, E., Tallal, P., 1991. Cerebral structure on magnetic resonance imaging in language – and learning – impaired children. Archives of Neurology 48, 539–545.

Johnston, R.B., Stark, R.E., Mellits, E.D., Tallal, P., 1981. Neurological status of language-impaired and normal children. Archives of Neurology 10, 159–163.

Jongmans, M., Henderson, S.E., de Vries, L., Dubowitz, L., 1993. Duration of periventricular densities in preterm infants and neurological outcome at 6 years of age. Archives of Disease in Childhood 69, 9–13.

Kaplan, E., 1968. Gestural Representation of Implement Usage: An Organismic Developmental Study. Ph.D Thesis, Clarke University.

Knuckey, N.W., Apsimon, T.T., Gubbay, S.S., 1983. Computerized axial tomography in clumsy children with developmental apraxia and agnosia. Brain and Development 5, 14–19.

Kools, J.A., Tweedie, D., 1975. Development of praxis in children. Perceptual and Motor Skills 40, 11–19.

Losse, A., Henderson, S.E., Elliman, D., Hall, D., Knight, E., Jongmans, M., 1991. Clumsiness in children – do they grow out of it? A 10-year follow-up study. Developmental Medicine and Child Neurology 33, 55–68.

McDonald, S., Tate, R.L., Rigby, J., 1994. Error types in ideomotor apraxia: A qualitative analysis. Brain and Cognition 25, 250–270.

Mozaz, M.J., 1992. Ideational and ideomotor apraxia: A qualitative analysis. Behavioural Neurology 5, 11–17.

Nicolson, R.I., Fawcett, A.J., 1994. Comparison of deficits in cognitive and motor skills among children with dyslexia. Annals of Dyslexia 44, 147–164.

Overton, W., Jackson, J., 1973. The representation of imagined objects in action sequences: A developmental study. Child Development 44, 309–314.

Powell, R.P., Bishop, D.V.M., 1992. Clumsiness and perceptual problems in children with specific language impairment. Developmental Medicine and Child Neurology 34, 755–765.

Raven, J.C., Court, J.H., Raven, J., 1986. Raven's Progressive Matrices and Raven's Coloured Matrices. Lewis, London.

Semel, E.M., Wiig, E.H., Secord, W., 1980. Clinical Evaluation of Language Fundamentals – Revised. Psychological Corporation, San Antonio, TX.

Tallal, P., Katz, W., 1989. Neuropsychological and neuroanatomical studies of developmental language/reading disorders: Recent advances. In: von Euler, C., Lennerstrand, G. (Eds.), Brain and Reading. Stockton Press, New York.

Wechsler, D., 1990. Wechsler Preschool and Primary Scale of Intelligence – Revised. Psychological Corporation, Sidcup.

Whitmont, S., Clark, C., 1996. Kinaesthetic acuity and fine motor skills in children with attention deficit hyperactivity disorder: A preliminary report. Developmental Medicine and Child Neurology 38, 1091–1098.