Language and executive functioning in children with ADHD
Parigger, E.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Research method

This chapter describes how the research questions, as set out in Section 3.4, will be answered. The selection of subjects is dealt with first (Section 4.1), followed by the procedure (Section 4.2), the various tasks (Sections 4.3 and 4.4), and the statistical analyses (Section 4.5).

4.1 Subject selection

In this section the characteristics of the children in the three research groups are discussed. The research groups with typically developing children and with SLI children serve as controls to the research group with ADHD children.

Section 4.1.1 describes the inclusion and exclusion criteria used for the selection of the group with typically developing children. Section 4.1.2 provides further specifics of the SLI group, and Section 4.1.3 describes further specifics of the ADHD group. A summary of the details is presented in Section 4.1.4. More specifically: age, gender, non-verbal intelligence scores and symptoms of inattention and hyperactivity will be discussed.

16 General characteristics of children with SLI and ADHD were already discussed in Sections 2.1.1 and 2.2.1 respectively.
4.1.1 Typically developing group

The typically developing children came from two state schools in the north of the Netherlands. The inclusion and exclusion criteria used are summarized in Table 4-1. The parents of the children received a background questionnaire to (double-)check these criteria17. It consisted of 29 questions, mostly multiple-choice, but also some open-ended items.

A total of 22 typically developing children participated in the study (see Table 4-2). More boys were selected than girls in order to make the gender split comparable to the ADHD group.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range 7;0-8;11</td>
<td>Physical/perceptual problems</td>
</tr>
<tr>
<td>Monolingual Dutch</td>
<td>Cognitive problems</td>
</tr>
<tr>
<td></td>
<td>Psychiatric problems</td>
</tr>
</tbody>
</table>

Table 4-1: inclusion and exclusion criteria for the typically developing children

First of all, the children had to be 7 or 8 years of age. This age range was selected to match the desired age range for the ADHD group, and was chosen for two reasons: (1) ADHD is often not diagnosed before the age of seven and (2) ADHD language problems are most prominent before the age of nine. Secondly, only monolingual children were selected. This was because bilingualism might affect the language measure comparisons in the study (Genesee, 2003; Goorhuis and Schaeerlaekens, 2000). Children with severe physical problems were excluded. In particular, problems in hearing and vision may affect language development (Mills, 1993; Mogford, 1993; Knoors, 2001; Baker, 2006). Cognitive problems may also affect language development (Chapman, 1995). Therefore, all children were tested for non-verbal intelligence (SON-R; Snijders, 2001).

17 This questionnaire can be obtained by making a personal request to the author (e.m.parigger@uva.nl).
Furthermore, language development can be hampered by the presence of psychiatric problems (Blankenstijn & Scheper, 2003). So, in addition to the background questionnaire, which asked about psychiatric problems in general, an extra questionnaire gathered information about externalizing behavior in particular (VVGK; Oosterlaan, Scheres, Antrop, Roeyers and Sergeant, 2000; also see Section 4.1.3).

Originally, the TD group consisted of 25 children. Three children were excluded after the start of the study: one child was later diagnosed with ADHD; another child turned out to be bilingual, and the third was excluded at the discrimination of the researcher.

4.1.2 SLI group

The SLI children in this study came from two schools for language-impaired children in the north of the Netherlands. The inclusion and exclusion criteria from Table 4-1 also applied to the SLI group. A total of 19 SLI children participated in the study (see Table 4-2). The gender division was matched to that of the ADHD group.

The diagnostic label of SLI does not officially exist in the Netherlands. The diagnosis used is called ‘ernstige spraak/taalmoeilijkheden’ (ESM; severe speech/language deficits). It is normally made on the basis of the guidelines drawn up by Resing et al. (2005). They distinguish four problem areas: (1) auditory processing problems, (2) speech production problems (dyspraxia), (3) grammatical problems and (4) lexical/semantic problems. ESM is diagnosed when a child performs at least 1.5 SD below the mean on two subtests for at least two of the above problem areas, or when a child performs more than 2 SD below the mean on a general language test. This corresponds to the severity criterion of, for example, Stark and Tallal (1981) and Tomblin et al. (1997).
The SLI diagnosis is, however, also based on exclusion criteria (see Section 2.1.1). The school files of the children were consulted to ensure that no other clinical conditions, including ADHD, were diagnosed. Furthermore, intelligence was assessed using a non-verbal test for children in the age range of 5;6 up to 17 years (SON-R; Snijders, et al., 1988). This test consists of seven subtests, but the short version that was used in this study has four subtests and has been shown to be as valid and reliable as the longer version (Tellegen and Laros, 2003). The cut-off for non-verbal intelligence was set at 80. Taking a cut-off of 80 is not uncommon, also because intelligence does not correlate strictly with language outcomes (Fey, Long and Cleave, 1994; Bishop, 1997). Moreover, SLI children as a group tend to have somewhat lower scores, suggesting that low average intelligence is a common component of the SLI profile (e.g. Plante, 1998)\(^{18}\).

Most studies in the international literature consider grammatical impairment to be the core problem area in SLI, either explicitly, or implicitly (see, for example, De Jong, 1999). For this reason, and in order to increase homogeneity in the SLI group, all children selected had to have at least grammatical problems. They were not excluded if they had auditory processing problems and/or lexical/semantic problems, but they were excluded if they had speech production problems (dyspraxia). The focus in this study is on language problems, not on speech problems. Furthermore, it would be difficult to interpret the results from children with speech problems on tests such as the non-word repetition test (see Section 4.3.2).

To summarize, when comparing the SLI group in this study to SLI groups in the international literature, this group mostly resembles ‘typical SLI’, described by Bishop (2004). As discussed in Section 2.1, she coined the term to refer to a distinctive subtype within the SLI group as a whole. This subtype is mainly defined by

---

\(^{18}\) A cut-off of 80 is still well above the cut-off of 70 for mental retardation.
morphological/syntactic problems (such as immature sentence structure and omission of grammatical morphemes), although phonological difficulties often occur as well.

The SLI group originally consisted of 23 children but four had to be excluded because their non-verbal intelligence score was below 80. Intelligence had to be tested during the course of the study, since testing beforehand was not possible (also see Section 4.2).

4.1.3 ADHD group
The general inclusion and exclusion criteria set out in Table 4-1 also applied to the ADHD group\(^\text{19}\). A total of 26 ADHD children participated in the study (see Table 4-2). Only ADHD children from the combined and from the mainly hyperactive-impulsive subtypes were included in the study. This is because the inattentive subtype may best be characterized as a distinct disorder (see review of Milich et al., 2001; also see Section 2.2.1). Inclusion of the inattentive subtype would make the ADHD group less homogeneous. ADHD is distributed unevenly over boys and girls in the general population, in a ratio of 80% to 20% (Gezondheidsraad, 2000). An effort was made to also reflect this distribution in the ADHD research group.

The ADHD children were recruited in two ways. They either came from an outpatient clinic for children and adolescents with psychiatric disorders, or they were found through an advertisement on the website and in the magazine of an association for parents of children with learning, developmental and/or behavioral disorders\(^\text{20}\).

The outpatient clinic has three departments in the north of the Netherlands and all participated in this study. Several steps are taken in this clinic to ensure valid ADHD diagnoses, the most important

---

\(^{19}\) One of the ADHD children just turned 9:0 when tested. It was decided to keep him in the research group, but to observe closely whether or not his scores deviated from the scores of other ADHD children. This proved not to be the case.

\(^{20}\) This parents’ association is called BALANS. Their website can be found at www.balansdigitaal.nl.
being that the results of the children’s individual reports (i.e. reports from intake and an anamnesis, a psychological report, a psychiatric report and observational reports from the home and school environment) are discussed in an interdisciplinary team consisting of psychiatrists, psychologists, and social workers. A diagnosis is only accepted if consensus can be reached.

Extra care had to be taken when considering ADHD children recruited via the parents’ association. There is considerable variation in diagnosing ADHD in the Netherlands, and although it is done carefully at the clinic described above, it is not always the case elsewhere. In principle, it is even possible that the family doctor diagnoses ADHD on the basis of a 10-minute consultation. The parents of the children in the study were therefore asked to fill in a short additional questionnaire to provide details of the diagnosis. Children were only included if they had also been seen by a paediatrician or a child psychologist/psychiatrist who had confirmed the ADHD diagnosis. Some of the children participating via the parents’ association came from areas outside the urban area where the other children were tested. However, these children were included since they spoke standard Dutch, both at school and at home.

To further check on the ADHD diagnoses, the parents of all children who had given consent, and the teachers of these children completed a questionnaire on behavioral symptoms. This questionnaire is intended for children in the age range of 6 to 12 years and consists of 42 items, relating to behavioral symptoms listed in the DSM-IV categories of ADHD, ODD and CD (VVGK; Oosterlaan et al., 2000). The psychometrics of the VVGK questionnaire are reasonably good (COTAN, 2004). The parents’ and/or the teacher’s

---

21 This questionnaire can be obtained upon application to the author (e.m.parigger@uva.nl).
22 The VVGK is a translation of the well-known questionnaire on disruptive behaviour disorders (DBD; Pelham, Gnagy, Greenslade and Milich, 1992).
scores on the scales hyperactivity-impulsivity/inattention had to fall in the (sub) clinical range for the ADHD children to be included.

From the original set of 28 ADHD children, two had to be excluded: one could not be tested for the second and third time due to illness; another because one of his parents mainly spoke English to him.

4.1.4 Overview of subjects

In Table 4-2, age in months and gender of the children in the ADHD group, the SLI group and the TD group are presented. Non-verbal intelligence (SON-R; see Section 4.1.2) and ADHD symptoms (VVGK; see Section 4.1.3) are also reported for these three groups.

<table>
<thead>
<tr>
<th></th>
<th>ADHD (n=26)</th>
<th>SLI (n=19)</th>
<th>TD (n=22)</th>
<th>Total (n=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months (mean and SD)</td>
<td>97.73 (6.32)</td>
<td>99.74 (5.60)</td>
<td>97.59 (5.80)</td>
<td>98.25 (5.93)</td>
</tr>
<tr>
<td>Gender–male (% and n)</td>
<td>81% (21)</td>
<td>79% (15)</td>
<td>73% (16)</td>
<td>78% (52)</td>
</tr>
<tr>
<td>Gender–female (% and n)</td>
<td>19% (5)</td>
<td>21% (4)</td>
<td>27% (6)</td>
<td>22% (15)</td>
</tr>
<tr>
<td>Non-verbal IQ (mean and SD)</td>
<td>108.65 (12.55)</td>
<td>98.00 (13.59)</td>
<td>112.95 (10.45)</td>
<td>107.04 (13.45)</td>
</tr>
<tr>
<td>ADHD symptoms (mean and SD)</td>
<td>35.15 (9.26)</td>
<td>13.71 (8.06)</td>
<td>8.18 (8.09)</td>
<td>20.74 (15.11)</td>
</tr>
</tbody>
</table>

Table 4-2: age in months, gender, non-verbal IQ (SON-R) and ADHD symptoms (VVGK; summed totals of scales hyperactivity-impulsivity and inattention) in ADHD, SLI and TD groups. Data on symptoms of ADHD are missing for 5 of the 19 children with SLI.

The three research groups were matched for age in months as far as possible. The mean age across all groups is 98.3 months (i.e. 8
years and 2 months) and there was no significant difference in age between the ADHD, SLI and TD groups, \( F(2, 64) = .827, p > .05 \).

There was a representative but uneven gender distribution in the ADHD group (see Sections 2.2.1 and 4.1.3), that is, 81% boys and 19% girls. The SLI and TD children were selected in order to reflect this distribution.

Due to time constraints, it was not possible to match the children for intelligence. There turned out to be significant differences in IQ scores, \( F(2, 64) = 8.00, p < .05 \). More specifically, the SLI group had a lower IQ than both the ADHD group (\( p = .015 \)) and the TD group (\( p = .001 \)). The ADHD group and the TD group did not differ from each other (\( p = .537 \)).

As expected, there also were significant differences between the summed hyperactivity-impulsivity and inattention scale scores of the ADHD, SLI and TD groups, \( F(2, 59) = 64.68, p < .05 \). More specifically, as might be expected, the ADHD group has more such symptoms than both the SLI group (\( p = .000 \)) and the TD group (\( p = .000 \)). Moreover, the scores of the ADHD children are in the (sub)clinical range. The SLI group does not differ from the TD group (\( p = .176 \)).

Oppositional defiant disorder (ODD) and conduct disorder (CD) frequently co-occur with ADHD (e.g. Biederman, Newcorn and Sprich, 1991). Reading problems are also very common among children with ADHD (Gilger, Pennington and DeFries, 1992). It was not possible to exclude all ADHD children with one or more of these conditions (also see Section 2.2). However, the children were screened for ODD, CD and RP. We did this not only for the ADHD children, but also for the SLI and TD children. ODD and CD were assessed using the behavioral questionnaire, VVGK, as discussed in

\[\text{Matching for intelligence would mean testing a large pool of SLI children, because, as a group, they tend to have lower intelligence scores (e.g. Plante, 1998; also see Section 4.1.2).}\]
Section 4.1.3. To assess RP, two standardized reading tasks were used: a real word task (RWT; Brus and Voeten, 1973) and a pseudo-word task (PWT; Van den Bos, Lutje Spelberg, Scheepstra and de Vries, 1994). Raw scores on both tasks were converted to standardized scores with a mean of 10, and a SD of 3. RP is deemed to be present when children score more than 1 SD below the mean on both tasks.

<table>
<thead>
<tr>
<th>ODD (% and n)</th>
<th>ADHD (n=26)</th>
<th>SLI (n=19)</th>
<th>TD (n=22)</th>
<th>Total (n=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>69% (18)</td>
<td>86% (12)</td>
<td>86% (19)</td>
<td>79% (49)</td>
</tr>
<tr>
<td>subclinical</td>
<td>12% (3)</td>
<td>7% (1)</td>
<td>14% (3)</td>
<td>11% (7)</td>
</tr>
<tr>
<td>clinical</td>
<td>19% (5)</td>
<td>7% (1)</td>
<td>0% (0)</td>
<td>10% (6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CD (% and n)</th>
<th>ADHD (n=26)</th>
<th>SLI (n=19)</th>
<th>TD (n=22)</th>
<th>Total (n=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>77% (20)</td>
<td>79% (11)</td>
<td>91% (20)</td>
<td>82% (51)</td>
</tr>
<tr>
<td>subclinical</td>
<td>19% (5)</td>
<td>7% (1)</td>
<td>4.5% (1)</td>
<td>11% (7)</td>
</tr>
<tr>
<td>clinical</td>
<td>4% (1)</td>
<td>14% (2)</td>
<td>4.5% (1)</td>
<td>7% (4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RP (% and n)</th>
<th>ADHD (n=26)</th>
<th>SLI (n=19)</th>
<th>TD (n=22)</th>
<th>Total (n=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>23% (6)</td>
<td>68% (13)</td>
<td>18% (4)</td>
<td>34% (23)</td>
</tr>
<tr>
<td>not present</td>
<td>77% (20)</td>
<td>32% (6)</td>
<td>82% (18)</td>
<td>66% (44)</td>
</tr>
</tbody>
</table>

Table 4-3: presence of co-morbid oppositional defiant disorder (ODD), conduct disorder (CD), and/or reading disorders (RP) in the ADHD, SLI and TD groups. Data on ODD and CD are missing for 5 of the 19 children with SLI.

Table 4-3 shows the results of the screening for co-morbid oppositional defiant disorder (ODD) and conduct disorder (CD), as well as the screening for co-morbid reading problems (RP) in the ADHD group, the SLI group and the TD group. Data on ODD and CD will not be used in further analyses, and are displayed solely for information. Data on RP, however, will be used, in particular in Chapter 5 and 8.
4.2 Procedure

The parents of the children received an information letter, consent forms and several questionnaires when their children were selected. Most of them returned the forms and questionnaires within the testing period. Three parents failed to respond after several reminders. All were parents of SLI children, which means that fewer data are available for the SLI group. Their consent forms were collected via the contact person at school or via the people that treated the children.

The information letter contained precise information about the project, and about what was expected. It also provided information about who to approach in case of questions or complaints. The parents signed the informed consent form, also signed by the researcher, indicating that they were willing to participate in the study. On another form, they indicated whether or not they wanted the school or the family doctor to be informed about the results, when necessary. Moreover, they also indicated whether or not their child’s individual results could be used for presentations and/or publications. The three questionnaires were the background questionnaire (see Section 4.1.1), the behavioral questionnaire (see Section 4.1.3) and the language questionnaire (see Section 4.3.3).

The children were tested three times within as short a period of time as possible. This was preferably within three weeks, but eight weeks was the maximum. The sessions took place during school hours, mostly in the mornings, and were conducted in a separate room with no other people present. Each session lasted 45-90 minutes, including some time for informal talk. Most of the children were tested by the researcher. She has a background in clinical psychology and psycho- and patholinguistics. Some of the typically developing children were tested by two interns with a background in clinical linguistics, but only after extensive training by the researcher. In the first session the intelligence test (see Section 4.1.1) was
administered, and in the other two sessions language and executive functioning were assessed (see Sections 4.3 and 4.4). The language assessment was recorded on a digital video camera (Panasonic NV-GS75). It included a narrative task, a non-word repetition task and a sentence imitation task, as well as the two reading tasks already described in Section 4.1.3. The neuropsychological assessment contained measures of inhibition, working memory, planning, flexibility and non-verbal fluency. The order of testing, between and within the sessions was counterbalanced. There was one exception to this rule; the language assessment always started with the narrative task. As requested when entering the project, the children did not use any sort of medication (for example methylphenidate) before and during the language and neuropsychological sessions.

This was verified with both the parents and the children themselves before starting the sessions. After each successful session, the children received a token and at the end of the testing period, they received a small present. They all indicated that they had enjoyed themselves. The parents of the children in all three groups received a short descriptive summary of the performance of their child. In particular, feedback was given on the intelligence results and some of the more general language results.

4.3 LANGUAGE ASSESSMENT

As we saw in Section 2.2.2, children with ADHD have some problems with grammar and more severe problems in pragmatics. These language domains in particular will be the focus of this study (also see Section 2.4).

A summary of all tasks and outcome measures in the language assessment is presented in Tables 4-4 and 4-5. In Sections 4.3.1, 4.3.2 and 4.3.3, these tasks will be described in more detail.
<table>
<thead>
<tr>
<th>Frog story (narration)</th>
<th>Outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>number of analyzable T-units, total number of utterances not related to the story, total number of non-analyzable utterances, mean length of utterance (in words), percentage of subordinate conjunctions, percentage of dysfluencies, percentage of direct speech</td>
</tr>
<tr>
<td>Grammar</td>
<td>percentage of morpho-syntactic errors, percentage of morphological errors, percentage of syntactic errors, percentage of clustered morpho-syntactic errors</td>
</tr>
<tr>
<td>Pragmatics</td>
<td>total number of plot elements, total number of setting elements, total number of initiating events, total number of internal responses, total number of search attempts, total number of goals</td>
</tr>
</tbody>
</table>

Table 4-4: overview of outcome measures, subdivided in general, grammatical and pragmatic measures, in narrative task (frog story) used in language assessment

<table>
<thead>
<tr>
<th>Language tasks and questionnaire</th>
<th>Outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIT</td>
<td>number of items correct (out of 20)</td>
</tr>
<tr>
<td>NWR</td>
<td>number of items correct (out of 16), percentage of phonemes repeated correctly</td>
</tr>
<tr>
<td>CCC-II-NL</td>
<td>scales (10x): speech production, syntax, semantics, coherence, inappropriate initiation, stereotyped language, use of context, non-verbal communication, social relations, interests composites (3x): general communication score, social interaction score, pragmatic score</td>
</tr>
</tbody>
</table>

Table 4-5: overview of outcome measures in repetition tasks (sentence imitation task [ZIT] and non-word repetition task [NWR]) and language questionnaire [CCC-II-NL] used in language assessment
4.3.1 Narrative task

The book *Frog, where are you?* (Mayer, 1969) is for children, and has no text but only pictures. It has been used widely in international research, mainly for studies looking at the development of narrative abilities (Berman and Slobin, 1994). The story is about a boy who searches, together with his dog, for his lost frog. He looks in several places, but is unable to find the frog. In the end, however, he does find the frog and takes him back home. The story has a clear protagonist, who faces a problem, tries to solve it, and finally is able to come up with a solution. Dutch children in the targeted age range are used to these kinds of stories, and they are also familiar with storytelling (Roelofs, 1998). The stories produced by the children can be considered examples of semi-spontaneous language use.

The children were asked to flip through all pictures in the book before starting their narration. This was done to give them an overall idea of what was coming in order not to overload working memory when actually performing the narrative task itself. The experimenter then labeled the boy, the dog and the frog, and pointed at them. She then asked the children to look at the pictures again in sequence and while so doing to tell the story in their own words. She pretended to not know the story and made it clear that she could not see the pictures in the book herself. Feedback was given as little as possible. To be sure that the story had been well understood by the children, some comprehension questions were asked afterwards. All children were able to answer these questions.

The narratives produced by the children were transcribed by the researcher or by one of her interns - all with a background in clinical linguistics. The transcriptions were divided in T-units, using CHAT (MacWhinney, 2000). A T-unit is an abbreviation of terminable unit and it is defined as one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it (Hunt,
The transcriptions were double-checked and disagreements were solved after discussion between the researcher and the interns (also see Section 4.5.3). T-units with unintelligible words or unfinished T-units were considered not to be analyzable and coded as such. Elliptical utterances were also considered unanalyzable. The remaining analyzable T-units were divided into utterances that were related to the frog story, and those not related to the frog story. The former utterances were of interest for the analyses described in the remainder of this section (henceforward referred to as ‘T-unit(s) of the narrative’). The analyses were done on a grammatical and on a pragmatic level as set out in Sections 2.4 and 3.4.

The grammatical analyses described below are based on analyses in the Dutch STAP-procedure (Van den Dungen and Verbeek, 1999). This procedure is used both for language-disordered children and for children whose language is developing normally. It is the only method available for Dutch children in the age range of the study. It is aimed at 4- to 8-year-old children and determines the language production level in spontaneous conversations ‘outside the here and now’. However, STAP proved to be perfectly applicable to the semi-spontaneous language use in the frog stories, which can obviously also be considered as ‘outside the here-and-now’ language use.

First of all, the mean number of dysfluencies in the T-units of the narrative was determined. Four categories of dysfluencies were distinguished: (1) false starts, (2) self-corrections, (3) repetitions and (4) mixed constructions. False starts were identified when a child restarts an utterance, in a way different from the previous start of that

---

24 Pause words (e.g. ‘uh’) were counted separately because they were not considered to be dysfluencies in the same sense as false starts, self-corrections, repetitions and mixed constructions. Rather, they seem to reflect something like ‘thinking time’.
utterance, see Example [1]. In a self-correction, a child replaces part of a word or more while expressing the utterance, without starting the utterance anew, as in Example [2]. A repetition, as shown in Example [3] is a literal recurrence of part of a word or more. Finally, in a mixed construction a child starts the utterance as if going to form one structure, but then completes the utterance with a different structure. Words often become redundant when this happens, as can be seen in Example [4]. All (categories of) dysfluencies were totalled and divided by the total amount of T-units of the narrative.

[1] de bij het jongetje keek in een hol
   'the bee the little-boy looked into a hole'
   (ADHD26, male, 7;4.14 - coded as false start)

[2] en die is zit in een pot
   'and he is sits in a jar'
   (TD18, male, 7;3.26 - coded as self correction)

[3] waar ben je waar ben je
   'where are you where are you'
   (SLI18, male, 8;7.1 - coded as repetition)

[4] het jongetje houdt met zijn hand houdt ie het raam vast
   'the little-boy holds with his hand holds he on to the window'
   (ADHD7, female, 7;4.13 - coded as mixed construction)

Secondly, the mean length of utterance (MLU) was computed in order to obtain a global estimate of syntactic complexity. MLU was not calculated in morphemes, but in words, which is more appropriate given the advanced age of the children in this study.

---

25 Examples [1] to [17] are taken from data from this study. Participant group and number, as well as gender and age are given between brackets. All Dutch examples are translated into English. In general, the erroneous part of an utterance will be indicated by underlining. An asterisk is used whenever a certain linguistic form seems accurate in English, but is not in Dutch.
(Wells, 1985). More specifically, for the MLU, all communicatively used words (i.e. not including dysfluencies) in the T-units of the narrative were added up and then divided by the total amount of T-units of the narrative. As an additional measure of syntactic complexity, all subordinate conjunctions were totalled and divided by the total amount of T-units of the narrative. The use of subordinate conjunctions is indicative of more advanced linguistic abilities.

Thirdly, a grammatical error-analysis was carried out. According to STAP, errors in the inflection or conjugation of a verb, a noun, or an adjective are specified as separate morphological errors (see Examples [5], [6] and [7] respectively). Other morphological errors were allocated to a category ‘other’.

[5] de hond en de jongetje slaap
‘the dog and the little-boy sleep’
(SLI1, male, 7;11.16 - coded as verb conjugation error)

[6] (...) die kikker samen met een vrouwtje en allemaal jonkietjes
‘(...) that frog together with a little-female and a lot of little-ones’
(TD3, female, 8;4.0 - coded as noun conjugation error)

[7] en dan nemen ze één kleine kikkertje mee
‘and then they take one little-frog with them’
(ADHD20, female, 8;8.11 - coded as adjective conjugation error)

Syntactic errors are interpreted as errors in the structure of the utterance, caused by deletion, insertion, substitution or inversion (see Examples [8], [9] and [10]. Other syntactic errors were coded as ‘other’.

[8] plotseling ___ de kikker weg
‘all of a sudden the frog ___ gone’
(SLI4, male, 7;11.14 - coded as deletion)
when it was night and the boy who slept the frog crawled (...)’
(ADHD19, female, 8;3.14 - coded as insertion)

because sits he in there with his head’
(TD7, female, 8;5.15 - coded as inversion)

The percentage of morphological and the percentage of syntactic errors will be calculated over all T-units of the narrative. The percentage of the combined morphological and syntactic errors will also be calculated over all T-units of the narrative. Clustering of grammatical errors in a single utterance will also be considered, that is the percentage of T-units of the narrative having more than one grammatical error (see also Blankenstijn and Schepers, 2003).

Figure 4-1: illustration of the causal network model by Trabasso and Rodkin (1994). S=setting; E=event; IR=internal response; G1=goal 1, get frog back; G2=goal 2, find frog; G3=goal 3= search frog in particular locations; A=search attempt; O=outcome (positive or negative)

The main pragmatic analysis in this study is the plot analysis. As indicated in Section 2.2.2, ADHD children have been shown to have striking problems with the organization and monitoring of their narratives. The plot analysis will show up such problems. It focuses on the causal relations between the narrative events in the frog story.
The causal network model by Trabasso and Rodkin (1994) is shown in Figure 4.1. It forms the basis for the plot analysis in this study.

Trabasso and Rodkin (1994) argue that the frog story follows a hierarchical goal plan with unanticipated goal failures, goal reinstatements and goal successes. The plan begins with a setting (S; the boy has a pet frog), which is followed by an event (E; loses it) that happens to a protagonist. These respectively enable and psychologically cause an internal reaction (IR; shows concern over loss). The reaction leads to a goal (G1; plan to get the frog back). This goal motivates a subordinate goal (G2; plan to find the frog) to obtain it. The subordinate goal (G2) in turn motivates another subordinate goal (G3; plan to search for the frog in particular locations) to obtain it. Together those three goals constitute a plan. The plan is carried out through actions motivated by the third goal in the hierarchy (G3). This goal motivates an initial attempt (A; looking or calling) that fails as indicated by a negative outcome (O-; frog is not found and animals other than the frog suddenly appear). This failure is monitored and psychologically causes the reinstatement of the second goal (G2) that is motivated by the first goal (G1) that controls the overall plan. The frog story repeats this cycle six times. Finally, the seventh time, an attempt leads to a successful outcome (O+) for the third goal. This outcome enables another successful outcome (O+) at the level of the second goal (G2). This outcome causes a successful outcome (O+) for the first goal (G1), thus completing the hierarchical goal plan.

There is an extensive international research tradition based on the causal network model by Trabasso and Rodkin (1994). Blankenstijn and Scheper (2003) and Roelofs (1998) added some features to this analysis model. The scoring system focuses on whether certain narrative events, the so-called planning components, are reported. Planning components are only scored when a complete state of affairs (i.e. subject with verb) is present. In other words, when a
certain planning component is not accompanied by a subject and a verb, that component is not scored. Furthermore, the planning components have to be related when looking at the corresponding page in the booklet and from the viewpoint of the boy, possibly in combination with the viewpoint of his dog.

<table>
<thead>
<tr>
<th>Category</th>
<th>Planning components</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>introduction boy, dog, frog</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>boy possesses frog</td>
<td>1</td>
</tr>
<tr>
<td>Initiating events</td>
<td>boy asleep</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>frog leaves jar</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>boy awakes</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>boy finds jar</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>jar is empty</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>frog is gone</td>
<td>3</td>
</tr>
<tr>
<td>Internal response</td>
<td>boy is sad</td>
<td>3</td>
</tr>
<tr>
<td>Search attempts</td>
<td>boy searches frog in room</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>boy calls frog out of window</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>boy searches frog outside</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>boy searches frog in hole in ground</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>boy searches frog in hole in tree</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>boy calls frog from rock</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>boy searches behind log</td>
<td>21</td>
</tr>
<tr>
<td>Outcome</td>
<td>boy finds the frog/a(nother) frog</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>boy retrieves the frog/a(nother) frog</td>
<td>24</td>
</tr>
<tr>
<td>Internal response</td>
<td>boy is happy</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 4-6: all possible planning components of the frog story, divided over five categories. The corresponding pages in the booklet are also mentioned.

A total of 19 planning components is distinguished and divided over five categories: (1) setting, (2) initiating events, (3) internal respon-
ses, (4) search attempts\textsuperscript{26} and (5) higher order goals. A complete list is shown in Table 4-6.

Examples [11], [12], [13], and [14] and [15] provide illustrations of each category from the frog story transcripts.

[11] er is een jongen met een hond en een kikker
‘there is a boy with a dog and a frog’
(ADHD9, male, 8;10.6 - coded as setting: introduction)

[12] en de kikker is weg
‘and the frog is gone’
(SLI13, male, 8;7.12 - coded as initiating event: frog gone)

[13] en het jongetje ging in de boom kijken waar de kikker was
‘and the little-boy went in the tree to look for the frog’
(SLI9, female, 8;5.2 - coded as search attempt: hole tree)

[14] Tom ging weer naar huis met de kikker
‘Tom went home again with the frog’
(TD19, male, 8;5.9 - coded as outcome: boy retrieves frog)

‘Jan has to laugh really loud’
(ADHD25, male, 8;10.5 - internal response: boy happy)

In each transcript the planning components were coded and counted with CLAN. The variables of interest are the total number of worded planning components and the total amount of worded planning components in each of the categories.

\textsuperscript{26} Within the category of search attempts, a further differentiation is made in the location of a search attempt, (room, hole in tree etc.), the goal of a search attempt (frog), the action of a search attempt (subject with verbs like look, search, etc.) and the outcome of a search attempt (failure or success). A complete search attempt is only scored when action, goal and location are mentioned. When, in addition, the outcome is also mentioned, this is called a GAO-unit.
Finally, all uses of direct speech in the frog stories were coded (see Section 2.2.2). A distinction was made between direct speech with mention of the speaker, see Example [16], and direct speech without mention of the speaker, see Example [17]. The former is more sophisticated than the latter, but probably takes more processing effort, thus potentially placing an extra burden on executive functioning. All direct speech variables were counted and then divided by the total amount of T-units of the narrative.

[16] hij ging op de rots staan en riep: /“kikker, kikker”
‘he went to stand on the rock and shouted: / “frog, frog”’
(ADHD22, male, 7;10.10 - coded as direct speech with mention speaker)

[17] “stoute hond”
‘naughty dog’
(TD22, male, 7;3.11 - coded as direct speech without mention of the speaker)

4.3.2 Repetition tasks
Two different repetition tasks were administered: the sentence imitation task and the non-word repetition task.

The Dutch sentence imitation task that was used, the ‘zinnen imitatie taak’ (ZIT), was developed by Kraan-aan de Wiel (1989). Norm groups are available for this task. The task includes 20 sentences of varying complexity (and four practice items). The sentences are four to 10 words long, with a mean of 7.6 words. An example is shown in [18].

[18] gaan wij vanmiddag naar het zwembad?
‘are we going to the pool this afternoon?’
The sentences were spoken by a female native speaker, pre-recorded and played back, only once, on a laptop computer with loudspeakers. The responses were recorded on a digital video camera and were transcribed after the testing session. Each sentence was scored either as ‘right’ or as ‘wrong’. Poor performance on the sentence imitation task reflects grammatical problems. All scores were double-checked, and inconsistencies were resolved by the two coders (the researcher and one of the interns with a background in clinical linguistics). Additionally, qualitative analyses were performed on the error categories of this sentence imitation task (Dijkhuizen, 2010). These results will be summarized (see Section 5.2).

The Dutch non-word repetition task was developed by De Bree (2007), based on Dollaghan and Campbell’s task (1998). It contains 16 non-words of two to five syllables in length that conform to the Dutch phonotactic system. For each syllable length there are four items and the task starts with three practice items. The items do not contain any consonant clusters. Examples are so:taif and be:putamu:f. The non-words were pre-recorded by a female native speaker. They were played back on a laptop computer with loudspeakers. Repeated presentations of the items were not allowed. The responses of the subjects were recorded on a digital video camera and were transcribed after the testing session. The responses were scored as either ‘right’ or ‘wrong’. The percentage of phonemes repeated correctly for each word type (2-5 syllable words) was also calculated. Poor performance on the non-word repetition task reflects ‘poor encoding, storage and retrieval of phonological representations (De Bree, 2007: 112)’. All scores were double-checked, and inconsist-

27 The relation between scores on these verbal repetition tasks on the one hand and non-verbal working memory on the other hand will be discussed in Section 7.1.1.
encies were resolved by the two coders (the researcher and one of the interns with a background in clinical linguistics).

4.3.3 Language questionnaire

The children’s communication checklist-II, developed by Bishop (2003), was used as a general language measure. This questionnaire covers the child’s language use, focusing in particular on pragmatics (the main level of analysis in this study). The primary goal of this questionnaire is to obtain information on a broad range of pragmatic variables that supplements the more in-depth pragmatic analyses of the semi-spontaneous narrative samples (see Section 4.3.1). An additional advantage is that it offers a different perspective on the pragmatic skills of the children, since it is filled in by their parents. Information on various other aspects of language (for example phonology, syntax and semantics) can also be collected with the help of this questionnaire. A standardized Dutch translation of the questionnaire, produced by Geurts (CCC-II-NL; 2007), was available for children from 4 to 15 years of age. It consists of 70 items, divided into statements about possible weaknesses and statements about possible strengths of the child. Examples are given in [19] en [20]. The psychometrics of the CCC-II-NL generally are satisfactory (COTAN, 2004).

[19] lacht op gepaste momenten wanneer hij/zij met anderen praat

‘laughs at appropriate moments when he/she talks with others’

[20] stelt een vraag hoewel hij/zij het antwoord al heeft gekregen

‘asks a question although he/she has already received the answer’

The questionnaire has a multiple-choice format, with the possibility of choosing from four options (corresponding to the raw scores 0, 1, 2 and 3 respectively): ‘less than once a week; never’, ‘at least once
a week, but not every day’, ‘once or twice each day’ and ‘several times each day (or more than two times a day); always’. The items are divided over 10 scales: (1) speech production, (2) syntax, (3) semantics, (4) coherence, (5) inappropriate initiation, (6) stereotyped language, (7) use of context, (8) non-verbal communication, (9) social relations and (10) interests. Each scale has five positively worded items and two negatively worded items, resulting in a minimal raw score of 0 and a maximal raw score of 21. From these scales three composite scores can be derived: (1) the general communication score, based on scales 1 to 8 (2) the social interaction score, based on scales 1-5 and scales 8-10 and (3) the pragmatic score, based on scales 5-8. Normally, a general communication score above 104, which equals the 90th percentile, is taken as indicative of an SLI diagnosis. The interpretation of the general communication composite and the pragmatic composite is quite straightforward; higher scores mean more problems. However, this is not the case for the social interaction composite, where the interpretation is altogether different. This score is used to classify children, predominantly in research contexts. Children with a low negative social interaction score will mainly have structural language problems (as in SLI). Children with a high positive social interaction score will mainly have pragmatic problems (as in autism spectrum disorders).

4.4 NEUROPSYCHOLOGICAL ASSESSMENT

As set out in Sections 3.1 and 3.2, Pennington and Ozonoff (1996) discriminate between five executive functions: inhibition, working memory, planning, cognitive flexibility and non-verbal fluency. These executive functions will be tested in this study. The focus will be on non-verbal abilities in order to minimize the confound with linguistic abilities.
In order to standardize measurement of executive functioning as much as possible, we chose to use an automated neuropsychological testing battery, the CANTAB testing battery (Cambridge Cognition Limited, 2006)\textsuperscript{28}. The CANTAB battery was selected since it appeared more valid and reliable than other batteries available at the time of the study. CANTAB has been standardized on a large population aged between 4 and 90 in various studies (Strauss, Sherman and Spreen, 2006). Good levels of test-retest reliability have been reported (e.g. Fowler, Saling, Conway, Semple and Louis, 2002). Its validity has been established in a wide range of clinical populations (e.g. Fray, Robbins and Sahakian, 1996). More specifically, Williams et al. (2000) looked at SLI and hyperactivity, using the CANTAB. They included four groups of 6-year-old children: children with SLI, children with hyperactivity, children with SLI and hyperactivity and typically developing control children. SLI was not associated with reduced performance on any of the neuropsychological measures. Hyperactivity was associated with deficits on a test of attentional set shifting. They also had reduced spatial spans on a test of spatial working memory. There were no interactions. These results should be interpreted with caution however, because the children with hyperactivity were not officially diagnosed with ADHD. Moreover, the research groups were quite small.

From the CANTAB testing battery, six tests were selected. Two of these were control tests\textsuperscript{29}, passed by all children. Four measured executive functioning, in particular: inhibition (SST), spatial working memory (SWM), planning (SOC) and cognitive flexibility (IED - also

---

\textsuperscript{28} The CANTAB bibliography (see http://www.cantabbibliography.co.uk) contains over 700 peer-reviewed articles, all making use of CANTAB.

\textsuperscript{29} The MOT (motor screening; relevant as a control test for all other tests) and the BLC (big/little circle; relevant as a control test for the IED).
The specific tests, with their outcome measures, will be described separately in the remainder of this section.

<table>
<thead>
<tr>
<th>Function</th>
<th>Test</th>
<th>Outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibition</td>
<td>CANTAB-SST (stop signal task)</td>
<td>• SSRT</td>
</tr>
<tr>
<td>Working memory</td>
<td>CANTAB-SWM (spatial working memory)</td>
<td>• between-search errors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• within-search errors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• strategy score</td>
</tr>
<tr>
<td>Planning</td>
<td>CANTAB-SOC (stockings of Cambridge)</td>
<td>• problems solved in minimum moves</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• mean moves for n-move problems</td>
</tr>
<tr>
<td>Cognitive flexibility</td>
<td>CANTAB-IED (intra-/extradimensional shift)</td>
<td>• stages completed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• total trials (adjusted)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• errors blocks 2/5/7/9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• errors blocks 6/8</td>
</tr>
<tr>
<td>Non-verbal fluency</td>
<td>Paper-and-pencil-FPT (five point test)</td>
<td>• unique designs (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• perseverations (%)</td>
</tr>
</tbody>
</table>

Table 4.7: tests and outcome measures used in the neuropsychological assessment

The stop signal task (SST; clinical mode) is a classic stop signal task, which uses staircase functions to generate an estimate of stop

30 A test for fluency is not available in the CANTAB testing battery. This executive function is thus tested with a paper-and-pencil test, and will be described later on in this section.
31 It would have been informative to complement the tasks from this battery with the behavior rating inventory of executive function (BRIEF; Huizinga and Smidts, 2011). However, in the preparation phase of this study, this questionnaire was not yet available.
signal reaction time. The test gives a measure of a subject’s ability to inhibit a prepotent response (Aron, Dowson, Sahakian and Robbins, 2003). It takes around 20 minutes to administer. The task screen for the SST shows a white ring displayed to alert the subject, after which an arrow pointing to either the left or the right is displayed (see Figure 4-2). The test consists of two parts. In the first part, the subject is introduced to the press pad, and told to press the left hand button when a left-pointing arrow is shown and the right hand button when a right-pointing arrow is shown. In the second part, the subject is told to continue pressing the buttons as before, but to withhold the response, i.e. not press the button, when a beep is heard. This second part of the SST contains five blocks. At the end of each block, a feedback screen is displayed showing a graphical representation of the subject’s performance, which the administrator then explains to the subject, as well as encouraging them to go faster.

One SST outcome measure was selected for this study: stop signal reaction time (SSRT – last half). SSRT is an estimate of the length of time between the go stimulus and the stop stimulus at which the subject is able to successfully inhibit their response on 50% of trials. SSRT is calculated with the help of the stop signal delay (SSD) and the median reaction time on go trials (MRT-GT). The MRT-GT is self-explanatory. The SSD is calculated on the basis of the stop trials. In the stop trials, the auditory tone is played after the stop signal delay period. The timing of the auditory stop signal changes throughout the test. That is,

Figure 4-2: screenshot of the SST

Actually, the clinical mode was used for all CANTAB tasks that were selected for this study. The clinical mode is used when running the task only once per subject, and can be opposed to parallel mode, if available, for repeated testing.
sometimes the stop signal delay period is shorter, and sometimes it is longer. This depends on the subject’s past performance. Stopping should only be possible approximately 50% of the time. The dependent measure was calculated on the last half of the assessed sub-blocks in the test.

**Spatial working memory (SWM; clinical mode)** is a test of the subject’s ability to retain spatial information and to manipulate remembered items in working memory. It is a self-ordered task, which also assesses the use of a heuristic strategy (Owen, Downes, Sahakian, Polkey and Robbins, 1990). It takes about eight minutes to administer. The aim of the test is to fill up an empty column on the right hand side of the screen with blue tokens found, by using a process of elimination, in colored squares (boxes). The test begins with a number of boxes being shown on the screen (see Figure 4-3). The subject is instructed to search for blue tokens within these boxes. Touching a box reveals what is inside. Once a blue token has been found within a particular box, that box will never be used again to hide a token. Searching goes on until all blue tokens have been found. The number of boxes to be searched on the screen gradually increases over the trials, from 3 to 4 to 6 to 8 boxes. The color and position of the boxes used are changed from trial to trial to discourage the use of stereotyped search strategies.

![Figure 4-3: screenshot of the SWM](image)

Three SWM outcome measures were selected for this study: two error-scores and one strategy-score. Returning to an empty box where a blue token has already been found is referred to as a between-search error. A within-search error refers to responses to a box previously opened and shown to be empty. The strategy score is
based on the number of searches that start from the same location. A low strategy score signals better performance and is given to search sequences that consistently start from the same box. The dependent measures were calculated on 3-, 4-, 6- and 8-box problems.

The SWM and the sentence imitation/non-word repetition tasks (ZIT and NWR, see Section 4.3.2) all involve memory, although the former is non-verbal. It could therefore be the case that scores on the NWR will correlate with scores on the SWM (see Section 7.1.2).

Stockings of Cambridge (SOC; clinical mode) is a test of spatial planning (Morris, Downes, Sahakian, Evenden, Heald and Robbins, 1988). It takes around 10 minutes to administer. The subject is shown two displays containing three colored balls (see Figure 4.4). The subject must use the balls in the lower display to copy the pattern shown in the upper display. The balls can be moved one at a time by touching the required ball, then touching the position to which it should be moved. At first it is only necessary to move one ball, but later on, up to four balls have to be moved. Afterwards, a procedure controlling for motor performance is inserted: the upper display moves one ball at a time, repeating the moves made by the subject in the corresponding previous planning phase and the subject must follow the upper display by moving the balls in the lower display. A second block of planning problems of two, four and five moves then follows. The test is completed with a second block of matching motor control problems.

Two SOC outcome measures were selected for this study: problems solved in minimum moves and mean moves for n-move problems. The first one is a fundamental measure, recording the number of
occasions upon which the subject has successfully completed a test problem in the minimum possible number of moves. This is a measure of overall planning accuracy. The second measure describes the mean number of moves required by the subject to solve problems with solutions possible in two, three, four or five moves.

**Intra-/extradimensional shift (IED; clinical mode)** is a test of rule acquisition and reversal and as such, also reflects cognitive flexibility (Downes, Roberts, Sahakian, Evenden, Morris and Robbins, 1989). It takes around seven minutes to administer. It is a computerized analogue to the Wisconsin Card Sorting Test. The computer screen shows four squares, two of them empty, two of them filled with stimuli. These are either simple stimuli, made up of just one dimension (color-filled shapes), or compound stimuli, made up of two dimensions (color-filled shapes and white lines; see Figure 4-5). The subject starts by looking at simple color-filled shapes and must learn which one is correct by touching one of them. Feedback teaches the subject which one is correct, and after six correct responses, the stimuli and/or rules are changed. These shifts are initially intra-dimensional (color filled shapes remain the relevant dimension), then later extra-dimensional (white lines become the relevant dimension). There are nine stages to be completed. Subjects completing all nine stages are considered to have ‘passed the test’.

Four IED outcome measures were selected for this study: two scores for the number of stages/trials completed and two error-scores. The stages completed score consists of the total number of stages completed successfully. The total trials (adjusted) score consists of the total number of trials completed on all attempted stages, with an adjustment for any stages
not reached. Furthermore, one error-score reflects the total number of errors made in blocks 2, 5, 7 and 9. These errors provide a good measure of reversal learning. The other error-score reflects the total number of errors made in blocks 6 and 8. These errors are a measure of attentional flexibility.

As well as the neuropsychological assessment with the CANTAB testing battery, a traditional paper and pencil test was conducted in order to obtain a measure of non-verbal fluency. The five point test (FPT), designed by Regard (1982), with a three-minute-limit (Lee, Strauss, Loring, McCloskey and Haworth, 1997) and adapted for Dutch by Parigger (2006), measures non-verbal fluency. The FPT provides the subject with a sheet of paper showing eight rows of five squares. Within each of these squares, five points are depicted, just like the eyes on a dice. The purpose of the test is to make as many designs as possible by connecting the dots with lines. Between two points, only one line is allowed and only straight lines are permitted. Furthermore, the subject is told to only draw unique designs, that is, not to repeat designs already made. Two scores are of interest in the FPT, both derived from the raw score: the percentage of unique designs, and the percentage of repeated designs (perseverations). The higher the percentage of unique designs and the lower the percentage of perseverations, the more fluent subjects are supposed to be in their non-verbal responses.

4.5 Statistics

This section will first give an overview of the data collected in this study (Section 4.5.1). Moreover, the statistical analyses that were used will be explained (Section 4.5.2). The section also addresses the issue of interrater reliability (Section 4.5.3).
4.5.1 *(Missing) data*

Table 4-8 gives an overview of all available data in the ADHD group, the SLI group and the TD group.

Of the possible 938 datasets (14 different tasks/tests from 67 different children), 921 datasets (98.2%) could actually be collected. Of the missing 17 datasets (1.8%), the largest part is due to parents of SLI children failing to return questionnaires (three background questionnaires; five CCC-2-NL questionnaires; five VVGK questionnaires).

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>ADHD ( (n=26) )</th>
<th>SLI ( (n=19) )</th>
<th>TD ( (n=22) )</th>
<th>Total ( (n=67) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background (general information)</td>
<td>26</td>
<td>16</td>
<td>22</td>
<td>64</td>
</tr>
<tr>
<td>CCC-2-NL ([pragmatic] language)</td>
<td>26</td>
<td>14</td>
<td>22</td>
<td>62</td>
</tr>
<tr>
<td>VVGK (externalizing behavior)</td>
<td>26</td>
<td>14</td>
<td>22</td>
<td>62</td>
</tr>
<tr>
<td>Language assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frog story (narration)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>NWR (non-word repetition)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>ZIT (sentence imitation)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>RWT/PWT ([non]word reading)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>Neuropsychological assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SON-R (intelligence)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>SST (inhibition)</td>
<td>23</td>
<td>19</td>
<td>22</td>
<td>64</td>
</tr>
<tr>
<td>SWM (working memory)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>SOC (planning)</td>
<td>25</td>
<td>19</td>
<td>22</td>
<td>66</td>
</tr>
<tr>
<td>IED (cognitive flexibility)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
<tr>
<td>FPT (non-verbal fluency)</td>
<td>26</td>
<td>19</td>
<td>22</td>
<td>67</td>
</tr>
</tbody>
</table>

Table 4-8: overview of data obtained from the ADHD, SLI and TD groups
Also, from the CANTAB data, one SOC result is missing, due to software errors, from an ADHD child. Furthermore, three SST results are missing, all from ADHD children, since they failed to complete the task. The statistical analyses in Chapters 5, 6 and 7 were thus sometimes performed on fewer than the maximum of 67 children, but never on less than 62 children.

4.5.2 Statistical analyses

All data were entered into and analyzed with the statistical program SPSS (see Field, 2005).

The data were checked as to whether they were normally distributed, and had homogeneous variances. If the data were not normally distributed, as indicated by a Kolmogorov-Smirnov test, parametric tests were nevertheless used, but, where possible, checked with a non-parametric test. In case of disagreement between the two statistical tests, only the non-parametric test result was reported. In case of unequal variances, as indicated by Levene's test, Welch $F$ was reported. Some further checks were made for specific statistical tests. These included Mauchly's test for sphericity when conducting a repeated measures ANOVA (when significant followed by the Greenhouse-Geisser correction), and Box's test for equality of covariance matrices when conducting a MANOVA.

Null-hypotheses are rejected or accepted, depending on the $p$-values of the particular statistical tests. In this study, type I (i.e., 'false positive') and type II (i.e., 'false negative') errors were considered to be equally undesirable. Therefore, a fairly standard $p$-value of .05 was adhered to. Furthermore, effect sizes were reported. In case of parametric testing, this usually was the $r_p^2$. In case of non-parametric testing, it usually was Cohens’ $d$. Cohen (1992) considers an effect size of .2 to be small, of .5 to be medium, and of .8 to be large.
The dependent measures were analyzed using ANOVAs with group (three levels) as the between subject factor. When for one task there was more than one dependent variable, MANOVAs were used instead of ANOVAs. We co-varied IQ scores, because the scores of the groups differed, and intelligence can potentially be related to the outcome measures that were used. When necessary, outliers were removed from the data. This was never done without mention. Furthermore, planned comparisons were applied. With planned comparisons, contrasts between groups can be tested statistically. In this study, simple (last) and difference contrasts were used. Explorative correlations were also carried out. We opted for partial correlations, to control for the differences in IQ-scores.

Additionally, within-group and individual comparisons were made. This was normally done to compare various outcome measures and in order to do so, scores were converted to z-scores, with a mean of 0 and a SD of 1.

In case of multiple comparisons, Holm’s correction was applied to make sure that the chance for a Type I error did not exceed the .05 level. In fact, Holm’s correction is very much like a Bonferroni correction, but somewhat less conservative. The $p$-values of the relevant tests are ordered from the smallest to the largest. The smallest $p$-value needs to be smaller or equal to $\alpha/k$ ($\alpha$ is .05 and $k$ is the number of tests). The next smallest $p$-value needs to be smaller than $\alpha/k-1$ and so on, until the corrected $p$-value becomes larger than .05 (Holm, 1979).

4.5.3 Interrater reliability

As is clear from Table 4-8, three different sets of data were collected in this study: (1) data from the questionnaires, (2) data from the

---

33 However, when testing four groups, the Kruskall-Wallis test was performed. This is the non-parametric counterpart of an ANOVA. See for example Section 5.7 where two ADHD groups (one with and one without language problems), the SLI group and the TD group are compared.
language assessment and (3) data from the neuropsychological assessment.

An interrater reliability estimate was not needed for the data gathered from the questionnaires. Answers to questionnaires are normally not subject to many interpretational problems. The researcher put the answers to the questions into the statistical program SPSS. This was double checked by one of the interns, who all had a background in clinical linguistics, in order to eliminate mistakes. An interrater reliability estimate was also not needed for data from the neuropsychological assessment, because the CANTAB battery automatically computes all outcome measures. This is a standardized procedure, so that mistakes cannot occur.

The data from the language assessment are susceptible to error. This was dealt with in two ways. The sentence imitation task, the non-word repetition task and the (non-)word reading tasks were coded both by the researcher and by one of the interns. All coding inconsistencies were then resolved. The frog story was transcribed by one of the interns and checked by another intern. The researcher herself checked the transcripts for a third time and eliminated a few last minor errors. The coding of the frog story transcripts was done by the researcher herself. An intern also coded a random 10% of the transcripts of each of the three groups to be able to determine the interrater reliability. Interrater reliability in this study was defined as the percentage of agreement between the two coders. In general, the interrater reliability over all coding categories was good. More specifically, agreement was 90% for the grammatical outcome measures and 89% for the pragmatic outcome measures.