A test for measuring syntactic development in young children

J. E.P.T. (Liesbeth) Schlichting and Henk C. Lutje Spelberg

Language Testing 2003; 20; 241
DOI: 10.1191/0265532203lt256oa

A test for measuring syntactic development in young children

J.E.P.T. (Liesbeth) Schlichting *Department of Phoniatrics, University of Utrecht; Effatha, Christian Institute for the Deaf, Zoetermeer* and
Henk C. Lutje Spelberg *Department of Special Education, University of Groningen, The Netherlands*

This article concerns linguistic, psychometric and clinical aspects of the Test for Sentence Development, a subtest of a new Dutch language production battery, the Schlichting Test voor Taalproductie (Schlichting Test for Language Production). The purpose of the Test for Sentence Development is to assess the syntactic knowledge in the language production of young Dutch children. The Test for Sentence Development aims to give an index of the level of grammatical skills of children from 1;8 to 6;3 and makes use of functional imitation as an elicitation technique. The advantage of the new test is that very young children can be assessed systematically, which was not possible previously. In the standardization study, the tests were administered to a representative sample of 1049 Dutch children. The reliability and validity results justify the use of the test in clinical settings and language research.

I Introduction

In the past few decades, language development has become a major issue in child developmental studies. The reasons for this interest are threefold. The first is the theoretical interest in a language disorder that is usually referred to as Specific Language Impairment (SLI). A number of studies have been carried out to gain insights into the nature of this disorder. Bishop (1992: 3) describes it as follows: 'Specific language impairment is diagnosed where there is a failure of normal language development that cannot be explained in terms of mental or physical handicap, hearing loss, emotional disorder or environmental deprivation.' The second reason for the increased interest in language development is a clinical one. Speech and language pathologists and psychologists claim that intervention is much more effective when given at an early age. As children mature, the neural plasticity of the

Address for correspondence: J.E.P.T. Schlichting, Department of Phoniatrics, University of Utrecht, Utrecht, The Netherlands; email: L.Schlichting@inter.nl.net

 10.1191/0265532203lt256oa

brain diminishes, which may imply smaller results in effectiveness of treatment (Lenneberg, 1967). The third reason for the increased interest is a political one. Taking part in a modern western society requires an adequate level of language ability. Children who do not have an acceptable command of the first language in their speech community have difficulties in following the school curriculum, and may consequently have insufficient schooling to enrol in the industrialized and computerized job market. However, governments are not prepared to fund language intervention when the child has not been diagnosed correctly. The decision on whether a child is eligible for intervention in financial terms often depends on the outcome of structured assessment (Nye and Montgomery, 1989). In Belgium, for example, the health care authorities provide speech therapy for a child only when that child has a percentile of 3 or less on at least one language test.

To ascertain that a child needs intervention requires assessment of the language development of the child, and for this assessment valid and reliable instruments are needed. It is generally held, however, that the language development of children up to the age of four years cannot be tested with sufficient reliability and validity. To investigate this paradoxical situation for the Netherlands, a language battery to assess language ability in young Dutch children was developed (see Figure 1). This battery contains a Dutch version of the Reynell Verbal Comprehension Scale A (Reynell, 1985) which is called Reynell Test voor Taalbegrip (Van Eldik *et al*., 1995), and a newly designed test battery for language production, the Schlichting Test voor Taalproductie (Schlichting Test for Language Production; Schlichting *et al*., 1995). The test battery for language production consists of four tests that can be used independently:

- the Test for Sentence Development, which is the subject of this article;
- a Vocabulary Test containing 61 items, mostly eliciting a response with a picture as the main stimulus;
- a Test for Auditory Memory, measuring the ability to repeat a series of one-syllable words in 15 items; and
- a Vocabulary Checklist containing 309 words and 10 sentences.

Measuring grammatical development is probably the area of greatest difficulty in the language assessment of young children (Crystal *et al*., 1976). One of the tests in the newly developed Schlichting Test for Language Production, the Test voor Zinsontwikkeling (Test for Sentence Development), was specifically designed to measure this area. Various features of this test are described in this article. The aim of the test is to reliably and validly measure the syntactic productive knowledge of Dutch children from 1;8 to 6;3. The reason for writing

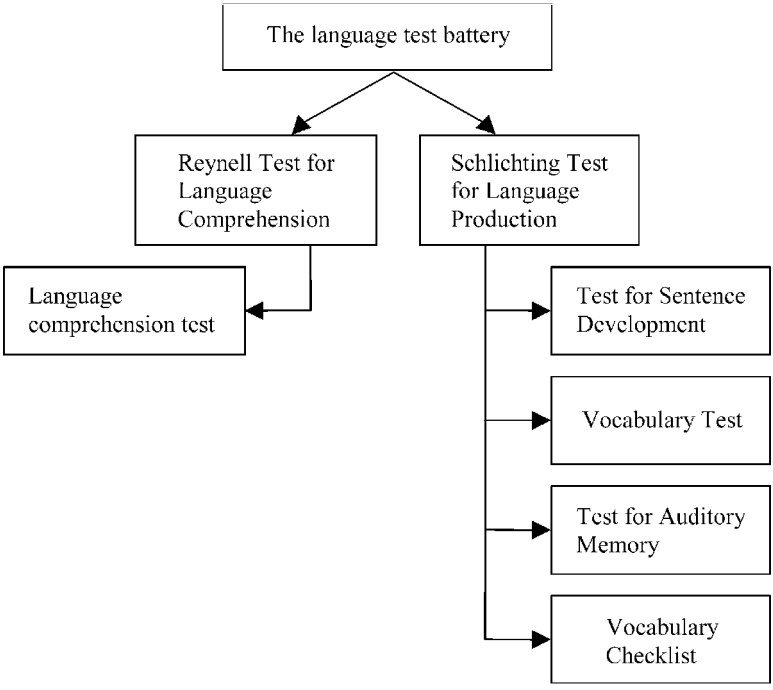

Figure 1 The subdivisions of the language test battery

this article is that, while in the Netherlands the test has been generally accepted in clinical settings and seems to present a possible solution to the problem of assessing syntax in very young children, apart from the test manual, no full description of the test has been published previously.

1 The assessment of syntactic development in young children

The assessment of syntactic development has seen many changes over the past decades. Syntactic development has been assessed primarily using the following techniques:

- analysis of spontaneous speech production;
- parents' reports; and
- elicitation procedures.

a Analysis of spontaneous speech production: Associated with an increasing theoretical interest in syntax in the 1970s, there was a renewed interest in syntactic development in children. It was found that children acquiring a certain language normally develop syntactic structures in a fairly similar sequence (Crystal *et al.*, 1976; Wells,

1985). This finding made it possible to describe the developmental sequence of sentence structures. For this description, it is necessary to have an index of syntactic ability. An early way of measuring syntactic ability which was described by Brown (1973) became widely used in a number of languages. In Brown's procedure, the data to be analysed are spontaneous utterances spoken in play or daily activities, including conversation. The child's utterances are transcribed and the mean length of utterance (MLU) is calculated, often in morphemes. Particular levels of MLU are associated with the child's syntactic productive development at a certain age. For example, a child with an MLU of 2.75 was expected to have (almost) mastered the use of the uncontractable copula (Brown, 1973: 271). The reliability and validity of MLU has been questioned extensively, however (Schlichting, 1996: 40). Children vary too much from one play session to another to be reliable in their sentence length. Wells (1985), who conducted an extensive study of British English children's spontaneous speech, also concluded that MLU was no longer predictable after the age of 42 months.

In speech pathology, the assessment and treatment of syntax (as well as morphology) has been an area which has been long neglected. Articulation and vocabulary were traditionally the prime objects for assessment and remediation. With a renewed interest in syntactic development in children, speech pathologists began to realize that syntax and morphology (to which we shall refer collectively as grammar) are important aspects of language disorders. This development naturally led to the need for assessment procedures and, in the wake of the theoretical interest in children's spontaneous utterances, assessment procedures were based on spontaneous speech. To this end, the child in the clinic is invited to play or converse with the speech pathologist or parent and the utterances produced by the child are transcribed. Subsequently, the utterances are analysed grammatically and entered in the various levels of a psycholinguistic profile. The result is a qualitative and quantitative description of the child's syntactic production, which can be used for assessment purposes and to set up goals for that child's intervention (Crystal *et al.*, 1976).

Various standard procedures for the analysis of spontaneous speech have been developed, e.g., American English (Lee, 1974) and British English (Crystal *et al.*'s Language Assessment Remediation and Screening Procedure (LARSP), 1976). The LARSP model was very successful in that it was used for the same purpose in various languages, e.g., Dutch TARSP (Schlichting, 1993). The drawback of the syntactic profiling is that while it is an extremely useful tool in syntactic research (Demuth, 1996), it is a time-consuming means of assessment. Taking a language sample within the context of a clinical

interview, transcribing the data and analysing the data can take three to six hours. This time factor is especially problematic when, in a clinical setting, a quick diagnosis of a child is required. An advantage of these procedures is that they do not only yield a level of syntactic development, but also provide detailed knowledge about the child's syntactic strengths and weaknesses, which can be used in planning intervention.

b Parents' reports: Parental report is widely used to gain insight into the language abilities of young children. Parents usually follow their children's language development with great interest, and have a good knowledge of their children's productive abilities, as opposed to the notion they have of their children's comprehension abilities. In Rescorla's (1989) Language Development Survey, the child's productive use of vocabulary is measured by means of a checklist for parents. Parents are presented with a list of 309 words, and are asked to mark which words their child uses spontaneously. A Dutch version of Rescorla's checklist for early vocabulary was included in the Schlichting Test for Language Production in 1995 (Schlichting *et al*., 1995; see Figure 1). The most widely used instrument in this area is the MacArthur Communicative Development Inventory (Bates *et al*., 1995), a set of two American English instruments for two age groups, 8–16 and 16–30 months, which now has versions in several languages (Fenson *et al*., 1993). Apart from the vocabulary, the MacArthur Communicative Development Inventory (CDI) also has items in the areas of sentences and grammar. Parents are asked to make an inventory of their children's knowledge about the use of word forms like *ate/ated*, and to select which of two sentences sounds most like the way their child talks at the present moment, for example *cookie mommy*/*cookie for mommy*. These checklists are much used as screening instruments; however, they were not meant to replace, and indeed they do not replace, full test procedures.

c Elicitation procedures: Elicitation procedures are communicative devices to encourage a child to produce a particular sentence. Some elicitation procedures require children to imitate a sentence, some to answer a question, others to correct a given sentence, etc. There is a sharp division between children from around four years old, who can be tested in a strict test procedure, and younger children, who do not easily conform to the demands of strict elicitation. Young children usually do not have the required linguistic awareness and the necessary task orientation to carry out the request to imitate a proffered sentence. We therefore set out to design a range of elicitation procedures by creating a situation in which very young children respond

in the required way without pressure, often even without asking them to respond. These elicitation procedures, which are used in the test, are the subject of the following section.

II Imitation as an elicitation procedure

1 Imitation as a theoretical basis for grammatical testing

When referring to test methods, Bachman (1990: 111) makes the following statement: 'One way to conceive of a language test is as a means for controlling the context in which language performance takes place.' The aim of the construction of the Test for Sentence Development was to create a context in which the syntactic structure in the input elicits a response from the test-taker which contains the same structure. In the present test, the controlling of the context is performed by presenting a stimulus that elicits a response of an imitative nature. Imitation is a concept with many definitions. In relation to the Test for Sentence Development, the definition of imitation adopted is: 'the phenomenon of words and structures from prior utterances appearing in following sentences' (Bohannon III and Stanowicz, 1989: 122).

The theoretical construct underlying the elicitation technique used in this test is syntactic activation. Syntactic activation or persistence is defined as a tendency to repeatedly employ the same syntactic form across successive utterances (Bock, 1986). For example, a speaker is more likely to produce a passive construction when, in the previous five utterances, a passive construction has occurred; the previous passive has acted as a primer or facilitator for the production of the passive used (Weiner and Labov, 1983).

Exact imitation and activation are closely related concepts. The difference between the two is that in exact imitation, be it partial or complete, the lexical items as well as the syntactic structures are identical, while this is not a necessary element in syntactically-activated sentences. For example, in the sentences below, if we consider (1) as the priming sentence or model, (2) has possibly been facilitated by (1), but it is not an exact or partial imitation. It is the result of an activation of the use of a passive construction.

1) The girl was seen in Amsterdam.
2) The woman was observed in Rotterdam.

Syntactic activation may thus be seen as imitation on a more abstract level. Imitation is a frequent phenomenon in child language. Bohannon and Stanowicz (1989) argue that 10% to 40% of the speech directed to children and 5% to 40% of children's speech is imitative, depending on the specific definition. These high percentages indicate

that imitation is a normal feature of discourse between parent and child.

An essential element of imitation in child, and perhaps adult, language is that young children cannot imitate utterances that contain structures that are not part of their linguistic systems. This view has been held for a long time (Ervin, 1964; Slobin, 1967; Slobin and Welsh, 1973). In the research of syntactic knowledge, elicited imitation has long been an important tool. Lust *et al.* (1996) argue that elicited imitation has been found to allow precise testing of children's knowledge of syntactic structures for research purposes. We suggest that elicited imitation, in a functional context, is also a possible vehicle for the assessment of syntactic structures in clinical settings. The hypothesis in using imitation in the Test for Sentence Development is that children can imitate utterances that are at, or slightly above, their level of linguistic ability in spontaneous speech.

2 Imitation in a functional context

Imitation in the present test is not the imitation of strings of words without meaning to the child. Young children are not motivated to imitate sentences that are not meaningful to them. Generally, in language testing imitated sentences do not serve a communicative function, and contextual cues are absent (McDade *et al.*, 1982). In the present test, the utterances are designed to have a communicative purpose and are presented to the child in a functional context.

Child language development is generally characterized by a moving away from the concrete and the here-and-now towards the abstract and the there-and-then. Very young children learn new words while handling concrete material (Rodgen *et al.*, 1977). Early syntactic structures are also closely linked to actions, be they the parents' or the child's actions. In the present test, we devised settings in which the examiner handles play material and at the same time presents an utterance pertaining to the handling of the material. The child is invited to imitate the examiner's action. While the child carries out the action, he or she also produces the utterance. For example, in item 13, the examiner introduces the choosing of pictures to be put in a paper frame. The examiner says, choosing a picture: 'I think I'll have the pear.' Then it is the child's turn to choose a picture and, without further stimulation, the child will typically produce the same utterance, in this case with a variation: 'I think I'll have the apple.' In choosing and taking the picture, the child focuses on the contents of the utterance, and because the child wants to imitate the action, he or she also copies the utterance. The production of this utterance has a linguistic function for the child in that it accompanies the action,

namely, the choosing and taking of the picture. The accompaniment of an action is a frequent function of utterances in child language (Wells, 1985: 233).

The role of intonation and other prosodic aspects are crucial in this way of eliciting syntactic structures. In children's early sentences, the intonational pattern is often the carrier of the structure and that characteristic is made use of in this test. The standardized use of prescribed intonational patterns is a requirement for the reliable use of the test. An audiotape containing the standard intonation patterns of the introductory sentences and the text of the stimuli is included in the test.

III Description of the instrument

1 The selection of the syntactic structures

The Test for Syntactic Development sets out to elicit utterances containing a selection of syntactic structures. These syntactic structures were meant to follow the normal sequence of acquisition in syntactic development with increasing complexity of the items. For this a detailed knowledge of syntactic development is necessary. For Dutch, enough knowledge was available with regard to children up to around four years of age (Schlichting, 1993; 1996). The development of sentence patterns, phrases and morphology commonly occurring in the spontaneous speech of Dutch children has been described in the Dutch version of LARSP (Schlichting, 1993; 1996). From the range of around 70 syntactic structures, 12 were selected to be incorporated in items representing the syntactic development of children up to 4. These 12 syntactic structures range from a non-specific one-word sentence to various complex sentences and some morphological items (see Appendix 1). Less is known about the syntactic development of Dutch children above the age of four years. Therefore, a study of syntactic development of children between 4 years and 6 years was carried out. Six children were followed in their syntactic development over a two-year period. In the first year, audio-tapes were made of the children's spontaneous speech at bimonthly intervals. During the second year, this process continued at three-monthly intervals. The children's speech was transcribed and screened for new syntactic structures. These structures led to the design of items 13 through 40.

Syntactic structures do not occur in a linguistic and cognitive void. They occur in sentences whose difficulty level for the child is determined by various factors. These factors may interact. If a particular sentence contains complex words, the child may have more difficulty with its syntactic structure (Crystal, 1987). Apart from the syntactic

complexity of a sentence, other factors – namely, the lexical items, the length of the sentence and the articulatory complexity – may contribute to the level of difficulty. The cognitive level also has to be taken into account. For reasons of content validity, the complexity level of all the characteristics of an item apart from the syntactic level was kept low. With regard to the selection of words, the general policy was to use only lexical items that are normally acquired at an earlier age than the syntactic structure in the test item. For example: the word *biggetje* ('piglet') was not used in the three-word sentence because at that stage children usually have not yet acquired that particular word. The cognitive level too was kept relatively low, so that the child's non-verbal cognitive level did not influence the correctness of the response. The length of a sentence cannot be seen as distinct from its syntactic complexity. A certain sentence length may be necessary for the expression of a certain message. Therefore, some quite long sentences were included. The longest item is 'This is a photo of a man with a bird on his hat' (item 39). It would be misleading to say that in the 40 items of the test, the children's knowledge of the syntactic structures is tested, but we can say that the knowledge of certain structures in certain linguistic (and cognitive) contexts is assessed.

2 Elicitation techniques

The test contains 40 items varying from one-word sentences to complex sentences. The material of the test consists of a folder with 40 drawings, for example, a field in which model animals are placed, or clowns who are going to put on various types of trousers, etc. For 13 of the items, especially the early ones, there are sets of concrete material like model animals, picture cards, buttons, etc. All items are introduced by a few sentences like: 'Here are two clowns' or 'Look, two fish'.

In the test, several elicitation techniques are applied. The main techniques are based on imitation in a functional context:

- exact imitation;
- imitation with variation; and
- other techniques.

For some structures, elicitation in an imitation task was not suitable, and for these items other techniques were used.

a Exact imitation: 17 items: In exact imitation, the child is asked to repeat the action and the words of the utterance precisely. This technique is especially suitable for the early items and is used in items

1–12. It is also suitable for five later items that concern the imitation of very long sentences.

> Example: item 4: subject + direct object + verb.
> The folder displays a field. In the middle of the field, a small dish is placed. On the examiner's side of the field, two model animals stand outside the field; on the side of the child, there are also two animals. The examiner introduces the item by telling the child that this is a story about animals. Then the examiner picks up one of the animals on his or her side and models the action of the animal eating from the dish whilst saying the target utterance the examiner wants to elicit: *die water drinken* ('that (one) water drink'). If necessary, the modelling of action and utterance are repeated with the examiner's second animal. Then, again if necessary, the examiner points to one of the child's animals. The child picks up an animal on his or her side, makes it drink, and produces the same utterance.

b Imitation with variation: 20 items: In imitation with variation, the child is asked to repeat the action and to imitate the utterance with the same syntactic structure, but with a change of one or more words.

> Example: item 25: the feminine singular personal pronoun.
> A picture with four children is presented (see Figure 2). The examiner points to the two boys, saying: 'he laughs, he cries.' The child points to the two girls in the picture, saying: 'she cries, she laughs.' The child has to change the personal pronoun and the order of the verbs.

c Other elicitation techniques: 3 items: One item asks a question and two items require a completion of a proffered sentence.

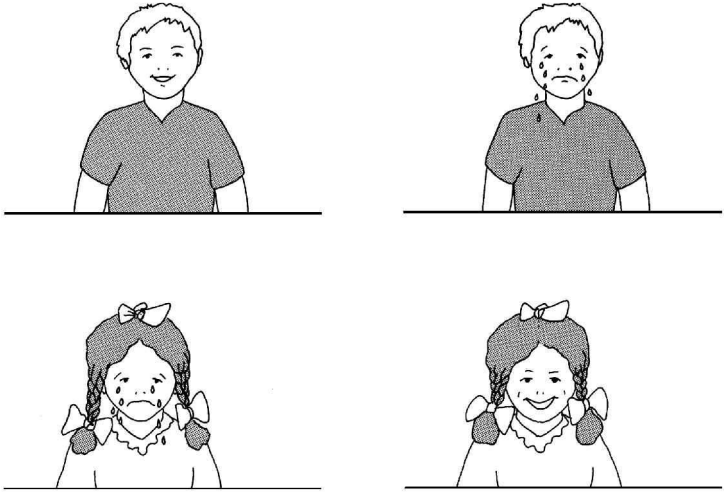

Figure 2 Pictures of item 25

3 Administration and scoring

The administration of the test takes less than 30 minutes. The 40 items are scored dichotomously: pass or fail. Passing an item results in one point given for a correct response. Correct responses are the target sentences with a number of variations which have been indicated in the manual together with a number of incorrect responses. Spontaneous utterances observed during assessment are alternative target responses for the first four items. For example, item 4 (see above and Appendix 1) is scored as correct if the child produces a three-word utterance containing a verb. After eight consecutive failures, the administration of the test must be stopped. Scoring a child's responses takes about 10 minutes.

In any test, the errors in the child's response may influence his or her performance, because of feelings of inadequacy. In the Test for Sentence Development, the child is asked for two responses, a non-target and a target response one. The child is always capable of giving the non-target response. Thus, the child will not feel the inadequacy usually associated with perceived failure. If the child cannot respond to the verbal stimulus as required, he or she can give the contents of the required response in a sentence with a lower syntactic level.

For example, in item 25, described above and in Figure 2, the non-target response is pointing out which girl is laughing and which girl is crying, and the child is keen to give this response. The target response is the use of the feminine pronoun. If the child does not automatically apply his or her knowledge about personal pronouns, the child selects a representation of the required response on a lower morphological level, viz. the male pronoun. In an item that requires a complex sentence, as in item 13, the child who cannot express the complex sentence, 'I think I'll have the apple' will give a simpler version: 'I'll have the apple.' The child will not produce the simpler structure if he or she can imitate the stimulus structure because, provided that the child has acquired that structure, he or she is naturally inclined to imitate the structure that has been activated in his or her linguistic system (see Section II.1 above).

IV Standardization, reliability and validity

The psychometric properties of the Test for Sentence Development were studied in a standardization study, conducted in the spring of 1993. This study served three purposes: establishing norms, estimating the reliability and studying the validity. The Test for Sentence Development was administered as part of a larger standardization study in which the other subtests of the Schlichting Test for Language

Production (Schlichting *et al.*, 1995) and the Reynell Test voor Taalbegrip (Van Eldik *et al.*, 1995) were standardized (see Figure 1). For the purpose of validation, several tests of language development, as well as tests for general and non-verbal cognitive development, were administered. In this section, the standardization sample, the results of standardization procedure, and reliability and validity studies are described.

1 Standardization sample

The age range of the children to whom this test battery was administered was 1;3 through 6;3. The reason for the lower bound of 1;3 was that the focus of the project was the assessment of very young children. The upper bound was set at 6;3 because, although there was no lack of tests for the age of 4 years and older, clinical experience had shown that these tests were not suitable for children between 4 and 6 years of age with substantial delays in development. The sample consisted of children with ages as close as possible, with a maximum deviation of two weeks, to 1;3, 1;9, 2;3, 2;9, 3;3, 3;9, 4;3, 4;9, 5;3, 5;9 and 6;3 years. The total number of participants was 1049, almost equally divided across the age groups and the two sexes (see Table 1). The sample was representative with regard to age, sex, region (north, west, east and south), density of population of the place of residence and socio-economic background (see Van Eldik, 1998: Chapter 3). For children under four years old, municipal authorities were contacted requesting addresses of children born in a certain month. For children over four, schools were randomly selected and requested to supply names of children whose parents agreed to their children's participation in the project. Only children whose parents were both native speakers of Dutch were included in the sample.

Table 1 Numbers of participants in the standardization sample by age and sex

Sex	Age groups											Total
	1;3	1;9	2;3	2;9	3;3	3;9	4;3	4;9	5;3	5;9	6;3	
Girls	48	47	48	46	49	48	40	46	48	46	48	514
Boys	47	51	53	49	47	48	46	50	47	49	48	535
Total	95	98	101	95	96	96	86	96	95	95	96	1049

2 Standardization procedure

The standardization procedure was carried out in two stages. In the first stage, the raw scores were transformed into normally distributed z-scores for each age group according to a method described by Lienert (1961: 339). In the second stage, all age groups were treated simultaneously. By means of stepwise regression analyses, the z-scores calculated in the first stage were estimated from both the raw scores and the exact age of the participants. This procedure yielded a formula to establish z-scores for intermediate age levels. This way of computing the norms led to norm tables with values per month. A second advantage of this method is that chance fluctuations caused by the relatively small size of the samples were, to some extent, eliminated. The z-scores were linearly transformed to a standard score called Sentence Quotient (SQ; ZQ in Dutch). The SQ has a normal distribution with an average of 100 and a standard deviation of 15. This scale was adopted because it resembles the widely used intelligence quotient, which has the same distribution. In clinical practice in the Netherlands, the comparison of scores on a language test to IQ scores is often used to establish that a child has a (specific) language impairment. Table 2 displays the means and standard deviations of the raw scores across the various age groups. In Figure 3, the raw score distributions of the age groups are shown graphically.

As can be seen from the mean raw scores, there is a steady increase in scores across the norm groups from age 1;9 to 5;9 (the age group of 1;3 year old has a mean score of lower than one, which is why this age group was not included in the norm tables of the published test; Schlichting *et al*., 1995). From 5;9 years to 6;3 years, the increase is only small: from 25.11 to 25.55.

3 Reliability

The reliability of the Test for Sentence Development was estimated in two ways. First, the lambda-2 coefficient (Guttman, 1945) was

Table 2 Means and standard deviations of raw scores in the various age groups

	1;3[a]	1;9	2;3	2;9	3;3	3;9	4;3	4;9	5;3	5;9	6;3
Mean	.36	1.79	4.93	7.45	10.69	14.26	16.81	19.07	22.58	25.11	25.55
Standard deviation	.52	1.51	2.65	3.40	5.22	6.26	4.49	5.69	5.40	5.14	4.64

Note: a This age group was excluded from the norm tables.

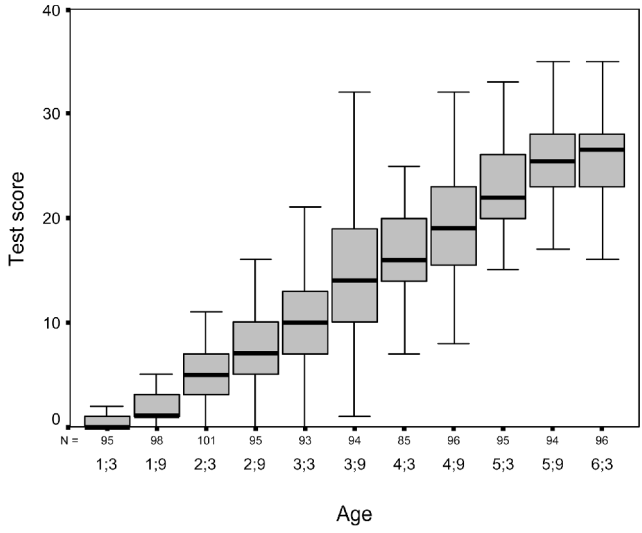

Figure 3 Boxplots of the distributions of the raw Sentence Development scores in the various age groups

Notes: The grey boxes indicate the area between the lower and upper quartile points, which means that 50% of the age groups fall in the grey areas. Except for around 20 outliers, all participants are in between the T-shaped ends of the plots.

used in order to determine the internal consistency of the test; secondly, test–retest reliability was computed. According to Ten Berge and Zegers (1978) the lambda-2 coefficient is a better estimate of internal consistency than the frequently used Cronbach's alpha coefficient. The mean lambda-2 was .85. Table 3 shows the results per age group.

To obtain an indication of the test–retest reliability, 130 participants

Table 3 Reliability results per age group

Age group	Lambda-2
1;9	.79
2;3	.85
2;9	.88
3;3	.91
3;9	.92
4;3	.86
4;9	.88
5;3	.85
5;9	.82
6;3	.77

Note: The given coefficients are lambda-2 values which are explained in the text.

of different ages were tested twice, with a six-month interval. The test–retest reliability, computed as the Pearson product-moment correlation, was .75. In view of the length of the interval between initial test and retest, this outcome was considered to be satisfactory. The Netherlands Institute of Psychologists (NIP) judged the overall reliability to be 'good' (Evers *et al*., 2000). The qualification 'good' refers to a reliability value of at least 0.90 for tests on which important, irreversible decisions are based, and at least 0.80 for tests used for less important decisions, like referral to therapy or vocational guidance.

4 Validity

The main focus of our validity research was on construct validity. Regarding the content validity, we refer the reader to Section III.1 and Appendix 1, which show that the items contain syntactic structures that vary widely in complexity and length. Construct validity refers to the degree a test measures the construct to be measured, here syntactic development. This kind of validity is determined by testing predictions derived from theories of syntactic development. The relationship with age, gender and socio-economic background in this investigation provides indirect evidence for the construct validity, because syntactic development should increase with age, and be related to gender and socio-economic background. Direct evidence is obtained from the relationship with tests which measure language skills (convergent validity) and cognitive skills (discriminant validity), and from a study comparing children with specific language impairment. Another study pointing towards construct validity concerned language development in Turkish children acquiring Dutch as a second language.

a Age, gender and socio-economic background: Because the Test for Language Development is meant to be a developmental test, the mean raw scores should increase with age. Table 2 and Figure 3 show that the raw scores indeed answer this requirement. From 5;9 years the increase is rather small.

The rate of language development is known to be faster in girls than in boys (McCarthy, 1954; Fenson *et al*., 1993). If the Test for Sentence Development is a valid test, the difference between the two sexes should show up in the scores. The same holds for socio-economic status (SES), where children whose parents have a higher educational level have been shown to have higher language scores (Morisset *et al*., 1990). A two-way analysis of variance was performed to test the differences between the mean standard scores,

which are given in Table 4. Significant main effects were found for sex ($F(2, 849) = 13.050$, $p < 0.001$) and SES ($F(2, 849) = 16.962$, $p < 0.001$). With a mean standard score of 98.4, boys were shown to have a slower syntactic development than girls, whose mean standard score was 102.0. The SES levels show effects on syntactic development with mean standard scores of 96.4 for the Lower level, 99.1 for the Middle level and 104.1 for the Higher level. There was no significant interaction effect.

b Correlations with other tests: For the purpose of validation, several tests of language development, as well as tests for general and non-verbal cognitive development, were administered to the standardization sample. Pearson product-moment correlation coefficients between the Test for Sentence Development and the validation tests as well as the number of cases and the age range of the samples – are presented in Table 5. Correlations of the Test for Sentence Development with other language tests have to do with convergent validity. Such correlations should be higher than correlations with non-language tests.

The correlations with the language tests range from 0.63 to 0.72, while correlations with the non-language tests range from 0.42 – for a test designed to measure non-verbal intelligence – to 0.56. (Compare Skehan, 1998, who reported on correlations of 0.43 and 0.44 between IQ and language aptitude in adult learners.) Of special interest is the correlation between the Test for Sentence Development and the Test for Auditory Memory, 0.47. This correlation should not be too high because the performance on a test measuring syntactic development, using imitation, should not be determined mainly by auditory memory.

c Specific Language Impairment: One of the clinical uses of the Schlichting Test for Language Production is to identify children with

Table 4 Means and standard deviations of standard scores of the Test for Sentence Development for boys and girls at different levels of educational background

Educational background	Boys ($n = 422$)	Girls ($n = 433$)	Both sexes ($n = 855$)
Lower level ($n = 142$)	93.4 (13.0)	98.7 (13.7)	96.4 (13.5)
Middle level ($n = 435$)	97.2 (14.1)	100.9 (13.7)	99.1 (14.0)
Higher level ($n = 278$)	102.4 (15.3)	105.7 (14.7)	104.1 (15.1)
All levels	98.4 (14.6)	102.0 (14.2)	100.2 (14.5)

Table 5 Correlations (corrected for attenuation) between the Test for Sentence Development and the validation tests, as well as the age range and the number of cases

Validation test	Correlation	Age range[a]	*n*
RTT: Language comprehension[b]	.68	1;9–6;3	946
WO: Expressive vocabulary[c]	.72	1;9–6;3	939
AG: Auditory Memory[d]	.47	2;9–4;9	496
TvK: Language test for Children[e]	.63	3;9–6;3	91
TARSP: Spontaneous language[f]	.67	1;9–4;3	47
BOS: Bayley Scales: Mental Scale[g]	.47	1;9–2;3	54
GOS: Intelligence[h]	.56	2;3–4;3	58
RAKIT: Intelligence[i]	.57	3;9–6;3	96
SON: Nonverbal Intelligence[j]	.42	1;9–6;3	571

Notes:
a Age range of the children in the validation studies. Except for the TARSP score, all scores are age-independent standard scores.
b Reynell Test voor Taalbegrip (Van Eldik *et al*., 1995).
c Test voor Woordontwikkeling (Schlichting *et al*., 1995).
d Test voor Auditief Geheugen (Schlichting *et al*., 1995).
e Taaltests voor Kinderen (Van Bon, 1982).
f TARSP: Taal Analyse Remediëring en Screening Procedure (Schlichting, 1993).
g BOS 2–30: Bayley Ontwikkelingsschalen Mental Schaal (Van der Meulen and Smrkovský, 1983). The Dutch version of the Bayley Scales of Infant Development (BSID; Bayley, 1969).
h GOS: Groningse Ontwikkelingsschalen Cognitive Score (Neutel, Van der Meulen and Lutje Spelberg, 1996). The Dutch version of the Kaufman Assessment Battery for Children (K-ABC; Kaufman and Kaufman, 1983).
i RAKIT: Revisie Amsterdamse Kinder Intelligentietest (Bleichrodt, Resing, Drenth and Zaal, 1987).
j SON-R 2½–7: Snijders-Oomen Niet-verbale intelligentietest (Tellegen, Winkel, Wijnberg-Williams and Laros, 1998).

Specific Language Impairment (SLI). If the Test for Sentence Development is a valid test, children with SLI should have lower scores than children in the standardization sample. A group of 41 children with a clinical diagnosis of SLI was studied by Lutje Spelberg and Van Berk (1997). These children all attended a school for children with SLI or hearing impairment. Twenty-six out of the 41 Dutch SLI-children could be matched on age, sex and socio-economic background with children from the Dutch standardization sample. Their ages ranged from 4;3 to 6;3. Three language tests were administered: to measure language comprehension, the Reynell Test voor Taalbegrip (Van Eldik *et al*., 1995) was used; to measure language production, the Schlichting Test for Sentence Development and the Vocabulary Test (see Figure 1) were administered. The results indicated significant differences on all three tests between the two matched samples. In Table 6, the differences are expressed in the age independent quotients.

The results show that the SLI children's abilities are lowest in the area of syntactic development, with a mean standard score of 76.2.

Table 6 Means, standard deviations and t-test results of the differences between quotient scores on three language tests of matched SLI and standardization samples

Language test	Group		Student t-test results		
	SLI	Normal	T	df	Significance
Language comprehension[a]	86.3 (13.5)	98.9 (14.5)	–3.15	25	.004
Sentence development[b]	76.2 (8.7)	99.2 (8.9)	–9.56	25	< .001
Expressive vocabulary[b]	80.5 (12.1)	98.5 (13.7)	–5.91	25	< .001

Notes: a Measured with the Reynell Test voor Taalbegrip (Van Eldik *et al.*, 1995); b Subtest of the Schlichting Test voor Taalproductie (Schlichting *et al.*, 1995).

On the Expressive Vocabulary Test, their mean standard score was 80.5, while language comprehension was the area in which their lag was smallest with a mean standard score of 86.3 on the Reynell Test voor Taalbegrip. These results are a contribution to the construct validity of the language tests, and also suggest that the SLI diagnosis in this group of SLI children was based more on language production than on language comprehension skills.

d A pilot study in second language acquisition: Theories of syntactic development argue that the acquisition of first and second languages are different processes (Skehan, 1998). For children acquiring a particular language, the similarities in the acquisition of syntactic structures are substantial, regardless of whether this language is their first or second language (Dulay and Burt, 1975). For this reason, a pilot study in second language acquisition was set up. In the Netherlands, a growing percentage of children belongs to ethnic minorities, mainly from Turkey and Morocco. These children generally acquire Dutch as their second language. In educational settings, the assessment of the level of Dutch language acquisition is often required. For children from the age of five years in these minority groups special Dutch language tests are available, but these often prove too difficult for five-year-old children. In spite of the theoretical differences between first and second language acquisition, we hypothesized that the Test for Sentence Development might be a possible instrument to assess the language development of children in ethnic groups that are beginning to acquire Dutch as a second language. A pilot study was carried out with a group of 62 Turkish children acquiring Dutch as a second language (Marquering and Mateboer, 1993; Bolten and Klooster, 1994). No data were available as to their general intelligence or their skills in their first language, but they were not considered ‘at risk’ in terms of Specific Language Impairment. The

areas investigated were language comprehension and production. The instruments used were the Reynell Test for Language Comprehension and the Schlichting Test for Language Production, which were designed for first language acquisition. There were three age groups: 5;3, 5;9 and 6;3. Table 7 shows the results of mean raw scores with their standard deviations and quotient scores (for technical reasons the standard deviations of the quotient scores could not be computed).

Table 7 Means and standard deviations of raw scores, quotient scores on three language tests of Turkish children acquiring Dutch as a second language

Type of Score	5;3 years (n = 20)	5;9 years (n= 25)	6;3 years (n = 17)
Language comprehension[a]			
Quotient score	63	59	55
Raw score	42.4 (9.8)	47.9 (11.4)	54.4 (12.3)
Sentence development[b]			
Quotient score	75	72	67
Raw score	10.8 (2.8)	11.9 (2.9)	11.5 (2.5)
Expressive vocabulary[b]			
Quotient score	55	55	55
Raw score	10.3 (6.0)	15.6 (7.9)	16.5 (7.2)

Notes: a Measured with the Reynell Test voor Taalbegrip (Van Eldik *et al.*, 1995); b Subtest of the Schlichting Test voor Taalproductie (Schlichting *et al.*, 1995).

The mean scores of Language Comprehension, Sentence Development and Expressive Vocabulary were all significantly below the means of Dutch children (the mean quotients for Expressive Vocabulary all show the same quotient, 55, which is the lowest possible quotient score). The average mean quotient score across the three age groups for Sentence Development is 71.3. This result suggests that Turkish children acquiring Dutch as a second language have more problems with the acquisition of vocabulary than with grammar (compare Strating-Keurentjes, 2000: 29).

V Conclusions

The Test for Sentence Development has proved to be a valid instrument to measure the level of grammatical ability in Dutch children up to the age of 6;3. Its overall reliability is .85 and the test can be administered in under 30 minutes. Scoring a child's responses takes about 10 minutes. As such, the test is suitable for the assessment of grammatical ability in a clinical setting. An advantage of the test is that children have a positive attitude towards taking it, because they are focused on the content of the items rather than on the grammatical demands placed on them. Test administrators report that children like

handling the concrete material and often enjoy taking the test so much that they want to continue the test when, according to the stopping criterion, the administration should be discontinued. We suggest that an analysis of spontaneous language is used to determine more fully the targets of intervention if the Test for Sentence Development indicates that intervention is called for.

Imitation in a functional context is an adequate tool to measure grammatical ability. The communicative function of each item ensures that children are focused on the content of the communication, which provides a motivation for their cooperation. The items vary greatly in types of task, which again contributes to the children's motivation.

From Items 1 to 12, the increase in difficulty runs parallel with the grammatical knowledge of children, so that the test administrator can be confident that the subsequent items reflect a greater grammatical ability. This sequence of items could be constructed because the grammatical development of children up to four years of age shows a general pattern and was well known. In the test as a whole, eight consecutive failures is required before the administration of the test can be stopped. The reason for this is that after Item 12, the percentage of children passing an item varies to a great extent (see Appendix 1). In future research, a more gradual increase in the level of difficulty in the consecutive items may well result, so that a smaller number of failures suffices before the test administration is stopped.

From the age of 5;9 years to 6;3 years, the increase in mean raw scores is only from 25.11 to 25.55 (see Table 2). The question arises as to why the change between the two values is so small. The usual reason of a ceiling effect does not seem to apply because the distance between the means of the highest age group, 25.55, and the maximum score, 40, is three times the standard deviation. Moreover, no participant in the sample attains the maximum score of 40 as is shown in Figure 3. The assessment of syntactic development after 5 years of age is a research area which is very problematic. Kemper *et al.* (1995) evaluated various ways of assessing syntactic complexity in children's narratives. They studied six procedures assessing syntactic development in spontaneous language of children 5–10 years of age. According to three of these measurement procedures, there was little change in syntactic complexity from 5 to 10 years. The procedure that did show significant changes across the whole age range calculated the frequency of main and embedded clauses in various ways. This result suggests that the levelling out of the mean scores of the Test for Sentence Development between the age groups of 5;9 and 6;3 years of age might be caused by the fact that this test assesses the ability to reproduce various syntactic structures, while it is in the frequency

of use of syntactic structures in spontaneous language that the developing child between 5 and 6 shows progress. Further research with the age group over 6;3 will show whether this argument holds for the Test for Sentence Development.

A moderate correlation (0.47) was found between the scores on the Test for Sentence Development and the scores on the Auditory Memory Test (Van Eldik, 1998). The latter test measures auditory memory by asking the participants to repeat a number of well-known single-syllable words. This is one indication that auditory memory is not a substantial factor in imitation of utterances spoken in a functional context.

The clinical use of the test was shown in a study of Dutch children who attended a school for SLI children. The language abilities in this group of SLI children were seen to be lowest in the area of grammatical ability. This result is in agreement with the general observation that SLI children have more problems with the acquisition of grammar than with vocabulary (Crystal *et al*., 1976).

In a group of Turkish children without manifest language problems who were acquiring Dutch as a second language, assessment showed that the language level was lowest in the area of vocabulary, while the grammatical ability was developed to a higher level. We suggest that imitation in a functional context may prove to be a possible technique for measuring syntactic development in second language acquisition. It follows that if appropriate norms were obtained, the present test might also be used as a first step in the diagnostic process for the assessment of language impairment in children who are acquiring Dutch as a second language.

Test norms become obsolete and need to be checked periodically (Cronbach, 1961). In the Netherlands, the level of education of parents has risen considerably during the past two decades. The standardization of the present test battery took place in 1993. We suggest that between 2005 and 2010 new norms should be established.

Acknowledgement

The authors wish to thank the Dutch Organization for Scientific Research (NWO), who supported this research project (grant no. 575–256-040).

VI References

Bachman, L.F. 1990: *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bates, E., Dale P.S. and **Thal, D.** 1995: Individual differences and their

implications for theories of language development. In Fletcher, P. and Macwhinney B., editors, *The handbook of child language*. Oxford: Blackwell, 96–151.

Bayley, N. 1969: *Manual for the Bayley scales of infant development*. New York: The Psychological Corporation.

Bishop, D.V.M. 1992: The underlying nature of specific language disorders. *Journal of Child Psychology and Psychiatry* 33, 3–66.

Bleichrodt, N., Resing, W.C.M., Drenth, P.J.D. and **Zaal, J.N.** 1987: *Revisie Amsterdamse Kinder Intelligentietest* [Revised Amsterdam Child Intelligence Test]. Lisse: Swets and Zeitlinger.

Bock, J.K. 1986: Syntactic persistence in language production. *Cognitive Psychology* 18, 355–87.

Bohannon III, J.N. and **Stanowicz, L.** 1989: Bidirectional effects of imitation and repetition in conversation: a synthesis within a cognitive model. In Speidel, G.E. and Nelson, K.E., editors, *The many faces of imitation in language learning*. New York: Springer, 121–50.

Bolton, J. and **Klooster, C.** 1994: *De Reynell Taalontwikkelingsschaal bij Turkse kinderen* [The Reynell Scales for Language Development in Turkish children]. Groningen: Hanzehogeschool, Hogeschool van Groningen (unpublished paper).

Brown, R. 1973: *A first language: the early stages*. Cambridge MA: Harvard University Press.

Cronbach, L.J. 1961: *Essentials of psychological testing*. New York: Harper & Row.

Crystal, D. 1987: Towards a 'bucket' theory of language disability: taking account of interaction between linguistic levels. *Clinical Linguistics and Phonetics* 1, 7–22.

Crystal, D., Fletcher, P. and **Garman, M.** 1976: *The grammatical analysis of language disability. A procedure for assessment and remediation*. London: Edward Arnold/Whurr.

Demuth, K. 1996: Collecting spontaneous production data. In McDaniel, D., McKee, C. and Smith Cairns, H., editors, *Methods for assessing children's syntax*. Cambridge, MA: MIT Press, 3–22.

Dulay, H.C. and **Burt, M.K.** 1975: A new approach to discovering universal strategies of child second language acquisition. In Dato, D.P., editor, *Developmental psycholinguistics: theory and applications*. Washington, DC: Georgetown University Press, 209–33.

Ervin, S. 1964: Imitation and structural change in children's language. In Lenneberg, E., editor, *New directions in the study of language*. Cambridge, MA: MIT Press, 163–89.

Evers, A., Van Vliet-Mulder, J.C. and **Groot, C.J.** 2000: *Documentatie van tests en testresearch in Nederland* [Documentation of tests and test research in the Netherlands]. Assen: Van Gorcum.

Fenson, L., Dale, P.S., Reznick, S., Thal, D., Bates, E., Hartung, J.P., Pethick, S. and **Reilly, J.** 1993: *MacArthur Communicative Developmental Inventories*. San Diego, CA: Singular Publishers.

Guttman, L. 1945: A basis for analyzing test-retest reliability. *Psychometrika* 10, 255–82.

Kaufman, A.S. and **Kaufman, N.L.** 1983: *Kaufman assessment battery for children. Interpretive manual.* Circle Pines, MN: American Guidance Service.

Kemper, S., Rice, K. and **Chen, Y-J.** 1995: Complexities and growth curves for measuring grammatical development from five to ten. *First Language* 15, 151–66.

Lee, L. 1974: *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Lenneberg, E. 1967: *Biological foundations of language*. New York: John Wiley and Sons.

Lienert, G.A. 1961: *Testaufbau und Testanalyse*. Wertheim: Verlag Julius-Beltz.

Lust, B., Flynn, S. and **Foley, C.** 1996: What children know about what they say: elicited imitation as a research method for assessing children's syntax. In McDaniel, D., McKee, C. and Smith Cairns, H., editors, *Methods for assessing children's syntax*. Cambridge, MA: MIT Press, 55–76.

Lutje Spelberg, H.C. and **Van Berk, K.** 1997: Het taalniveau van kinderen met primaire taalontwikkelingsstoornissen [The language levels of SLI children]. *Stem-, spraak- en taalpathologie* 6, 101–11.

Marquering, P. and **Mateboer, F.** 1993: *Turks thuis, Nederlands op school* [Turkish at home, Dutch in school]. Unpublished manuscript, Hanzehogeschool, Hogeschool van Groningen, Groningen.

McCarthy, D. 1954: Language development in children. In Carmichael, L., editor, *Manual of child psychology*. New York: Wiley and Sons, 492–630.

McDade, H.L., Simpson, M.A. and **Lamb, D.E.** 1982: The use of elicited imitation as a measure of expressive grammar: a question of validity. *Journal of Speech and Hearing Disorders* 47, 19–24.

Morisset, C.M., Barnard, K.E., Greenberg, M.T., Booth, C.L. and **Spieker, S.J.** 1990: Environmental influences on early language development: the context of social risk. *Development and Psychopathology* 2, 127–49.

Neutel, R.J., Van der Meulen, B.F. and **Lutje Spelberg, H.C.** 1996: *GOS, Groningse Ontwikkelingsschalen* $2\frac{1}{2}$–$4\frac{1}{2}$ [Groningen Developmental Scales $2\frac{1}{2}$–$4\frac{1}{2}$]. Lisse: Swets & Zeitlinger.

Nye, C. and **Montgomery, J.K.** 1989: Identification criteria for language disordered children: a national survey. *Hearsay: The Journal of the Ohio Speech and Hearing Association* 4, 26–33.

Rescorla, L. 1989: The language development survey: a screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders* 54, 587–99.

Reynell, J.K. 1985: *Reynell developmental language scales, second revision*. Windsor: NFER-Nelson.

Rodgen, H.M., Jankowski, W. and **Alenskas, L.** 1977: A multifunctional approach to single word usage. *Journal of Child Language* 4, 23–44.

Schlichting, J.E.P.T. 1993: *TARSP: Taal Analyse Remediëring en Screening Procedure* [LARSP: Language Assessment Remediation and Screening Procedure]. 3rd edition. Lisse: Swets & Zeitlinger.

—— 1996: *Discovering syntax. An empirical study in Dutch language acquisition*. Nijmegen: Nijmegen University Press.

Schlichting, J.E.P.T., Van Eldik, M.C., Lutje Spelberg, H.C., Van der Meulen, S. and **Van der Meulen, B.F.** 1995: *Schlichting Test voor Taalproductie* [Schlichting Test for Language Production]. Lisse: Swets Test Services.

Skehan, P. 1998: *A cognitive approach to language learning*. Oxford: Oxford University Press.

Slobin, D.I. 1967: Imitation and grammatical development in children. In Endler, N., Boulter, S. and Osser, H., editors, *Contemporary issues in developmental psychology*. New York: Holt, Rinehart and Winston, 437–43.

Slobin, D. and Welsh, C. 1973: Elicited imitation as a research tool in developmental linguistics. In Ferguson, C. and Slobin, D., editors, *Studies in child language development*. New York: Holt, Rinehart, and Winston, 485–97.

Strating-Keurentjes, H.T. 2000: Lexicale ontwikkeling in het Nederlands van autochtone en allochtone kleuters [Lexical development in first and second language of Dutch kindergarten children]. Dissertation, Katholieke Universiteit Brabant, Tilburg.

Tellegen, P.J., Winkel, M., Wijnberg-Williams, B.J. and **Laros, J.A.** 1998: *SON-R 2½–7, Snijders-Oomen Niet-verbale intelligentietest* [SON-R 2½–7, Snijders-Oomen Nonverbal intelligence test]. Lisse: Swets & Zeitlinger.

Ten Berge, J.M.F. and **Zegers, F.E.** 1978: A series of lower bounds to the reliability of a test. *Psychometrika* 43, 575–79.

Van Bon, W.H.J. 1982: *Taaltests voor kinderen* [Language tests for children]. Lisse: Swets & Zeitlinger.

Van der Meulen, B.F. and **Smrkovský, M.** 1983: *Bayley Ontwikkelingsschalen BOS 2–30* [Bayley Development Scales BOS 2–30]. Lisse: Swets & Zeitlinger.

Van Eldik, M.C.M. 1998: *Meten van taalbegrip en taalproductie* [Measuring language comprehension and language production]. Groningen: Stichting Kinderstudies.

Van Eldik, M.C.M., Schlichting, J.E.P.T., Lutje Spelberg, H.C., Van der Meulen, S. and **Van der Meulen, B.F.** 1995: *Reynell Test voor Taalbegrip* [Reynell Test for Language Comprehension]. Lisse: Swets Test Services.

Weiner, E.J. and **Labov, W.** 1983: Constraints on the agentless passive. *Journal of Linguistics* 19, 29–58.

Wells, G. 1985: *Language development in the pre-school years*. Cambridge: Cambridge University Press.

Appendix 1 Test for Sentence Development: Item numbers, item popularities (percentages of participants passing the items) and syntactic structures included in the 40 items

Item	%	Syntactic structures of target utterances (phrases are mostly embedded in a clause)
1	91	Noun, or any spontaneous one-word sentence
2	84	Subject + Adverbial adjunct, or a spontaneous two-word sentence
3	81	Subject + Verb, or a spontaneous sentence containing a verb
4	72	Subject + Direct object/Adverbial adjunct + Verb, or a spontaneous sentence containing at least three words, including a verb
5	70	Subject + Auxiliary + Infinitive
6	71	Negation 'niet' (not)
7	70	Adverbial adjunct + Indefinite article + Noun
8	64	Pronominal adverb
9	60	Adverbial adjunct + Verb + Subject + Direct object (inverted order of subject and verb)
10	57	Subject + Verb + Direct object + two Adverbial adjuncts
11	52	Subject + Verb + three constituents
12	51	Main clause + Finite complement clause (Dependent question)
13	44	Main clause + Finite complement clause (Dependent statement)
14	49	Attributive negation + Noun
15	25	Main clause + Subject clause
16	19	Neuter indefinite article + Noun
17	15	Main clause + Nonfinite complement clause, including 'te' (to)
18	25	Adverbial clause introduced by 'omdat' (because)
19	23	Main clause + Nonfinite complement clause, including 'om te' (in order to)
20	21	Two main clauses connected by 'maar' (but)
21	39	Preposition in postnominal position
22	12	Noun phrase: Indefinite article + Adjective (with neuter suffix) + Noun
23	17	Reflexive pronoun + Verb
24	15	Noun phrase including 'zonder' (without)
25	12	Feminine personal pronoun (subject form)

26	13	Main clause including progressive form + Adverbial clause introduced by 'tot' (until)
27	33	Noun phrase: preposition + Neuter definite article + Adjective (+ Noun)
28	28	Main clause + Dependent question introduced by 'er' (there)
29	13	Feminine personal pronoun (non subject form)
30	14	Main clause + Relative clause
31	6	Main clause + Comparison clause
32	18	Present participle in an adjunct introducing a main clause
33	7	Mastery of subject-verb agreement
34	17	Past participle in inflected adjective position
35	5	Possessive pronoun, 3rd person singular, feminine.
36	4	Passive sentence including an agent adjunct
37	1	Possessive pronoun, 3rd person plural
38	9	Noun phrase: Indefinite article + Adverb + Adjective + Neuter noun
39	6	Noun phrase: Noun + Triple adjectival postnominal prepositional adjunct
40	9	Main clause + Direct object clause including 'er' (there) + compliment (in the superlative)