

Assessment of Sign Language Development: The Case of Deaf Children in the Netherlands

Daan Hermans

Royal Dutch Kentalis

Harry Knoors

Royal Dutch Kentalis And Behavioural Science Institute, Radboud University Nijmegen

Ludo Verhoeven

Behavioural Science Institute, Radboud University Nijmegen

In this article, we will describe the development of an assessment instrument for Sign Language of the Netherlands (SLN) for deaf children in bilingual education programs. The assessment instrument consists of nine computerized tests in which the receptive and expressive language skills of deaf children at different linguistic levels (phonology, vocabulary, morphosyntax, and narration) are assessed. We will describe how the instrument was developed and normed, and present some psychometric properties of the instrument.

In the 1980s and 1990s of the 20th century, schools for deaf children in the Netherlands started to adopt a bilingual view on the education of deaf children. Nowadays, most schools for deaf children in the Netherlands have developed and implemented a bilingual curriculum in their schools in which deaf children's acquisition of Sign language of the Netherlands (SLN) and spoken and written Dutch are both regarded as being essential for children's academic achievements (Knoors, 2007).

One of the implications of choosing a bilingual curriculum for deaf children is that their signing skills should be monitored carefully (Anderson & Reilly, 2002; Haug, 2005; Herman, 1998). This is especially the case for deaf children, as the population of deaf children is highly variable in sign language proficiency (Maller, Singleton, Supalla, & Wix, 1999; Mann, 2007). Like hearing children, deaf children of deaf

parents typically have acquired the grammar of a sign language in early childhood. However, the majority of deaf children are born to hearing parents, who often start to learn sign language themselves when the deafness of their child has been detected. At early ages, many deaf children of hearing parents will be delayed in the acquisition of their signing skills in comparison with deaf children of deaf parents (Boudreault & Mayberry, 2006; Herman & Roy, 2006; Hoiting, 2005; Maller et al., 1999). These children still have to acquire many aspects of the grammar of sign language in primary education. To allow for effective intervention for deaf children who struggle to acquire sign language, their signing skills should be carefully monitored (Haug, 2005). As schools for the deaf usually use sign language as the language for instruction in other school subjects (e.g., math instruction, literacy instruction), a delay in the acquisition of sign language will also affect the children's academic achievements.

Especially in the United States, work on the development of sign language assessment instruments for deaf children and adults started in the 1990s (Anderson & Reilly, 2002; Hoffmeister, 1999; Maller et al., 1999; Mounty, 1993, 1994; Prinz, Strong, & Kuntze, 1994). In Europe, tests assessing children's signing skills have now been developed for British Sign Language (BSL; Herman, Holmes, & Woll, 1999; Herman & Roy, 2006), Deutsche Gebärdensprache

No conflicts of interest were reported. Correspondence should be sent to Daan Hermans, Petrus Dondersplein 1, 5270 BA Sint Michielsgestel, PO Box, 7, The Netherlands (e-mail: d.hermans@pontem.nl).

(Fehrmann et al. [as cited in Haug, 2005]; Haug & Mann, 2005), and Langue de Signes Francais (Niederberger et al. [as cited in Haug, 2005]).

In the Netherlands, Jansma, Knoors, and Baker started the development of an assessment instrument for sign language in the Netherlands in 1995 as part of a larger project on bilingual education. Pilot versions of five SLN tests were developed within this project: two expressive vocabulary tasks, one receptive vocabulary task, and two receptive morphosyntactic tasks (Jansma, Knoors, & Baker, 1997). In 2001, we started a 5-year project that built upon the knowledge acquired in the project by Jansma et al. The aim of the project was to construct a standardized sign language assessment instrument for deaf children in primary education in the Netherlands.

Test Construction

One of the major problems in the development of morphosyntax tests in SLN is that there is little detailed descriptive research of SLN (see for similar problems in Australian Sign Language, Johnson, 2004; Schembri et al., 2002). In the 1980s of the past century, descriptive analyses of SLN started with the work by Schermer and colleagues (Harder & Schermer, 1986; Schermer, 1983, 1990; Schermer & Koolhof, 1990), and a first sketch of the grammar of SLN was described by Schermer, Fortgens, Harder, and de Nobel in 1991. Work on linguistic analyses of SLN has continued in (especially) PhD projects conducted at the University of Amsterdam, Leiden University, and Utrecht University (Bos, 1990, 1993; Coerts, 1992; Crasborn, 2001; Knoors, 1992; van der Kooij, 2002; van Gijn, 2004; Zwitserlood, 2003). The development of a sign language assessment for SLN was complicated as the work on the linguistic description of SLN was at the start of the project and still is in progress.

In the first 2 years of the project, the SLN tests were constructed in close collaboration with deaf and hearing coworkers of the Dutch Sign Language Centre in Bunnik and the Institute SLN in Utrecht. During this period, several SLN researchers and deaf informants were frequently asked for advice. The test battery consisted of nine computerized SLN tests. Each test assesses a deaf child's expressive and receptive skills at

Table 1 Overview of the Sign Language of the Netherlands tests

Linguistic level	Modality	Target age-group
Phonology	Receptive	4.0–8.0
Phonology	Expressive	4.0–8.0
Vocabulary	Receptive	4.0–12.0
Vocabulary	Expressive	4.0–10.0
Vocabulary	Expressive	8.0–12.0
Morphosyntax	Receptive	4.0–12.0
Morphosyntax	Expressive	4.0–12.0
Narration	Receptive	6.0–12.0
Narration	Expressive	6.0–12.0

a particular linguistic level: phonology, vocabulary, morphosyntax, and narration. Table 1 provides an overview of the SLN tests we constructed and administered during the norming and validation study.

All SLN tests were developed for deaf children in primary education, aged 4–12. As these children are in different stages of the acquisition of SLN, not every test was expected to be appropriate to an equal extent for every age-group. To illustrate, the tests assessing the receptive and expressive phonological skills of deaf children were mainly developed for the younger children (e.g., aged 4–8 years), as most of the older deaf children were likely to have already acquired the phonological system in SLN. Similarly, the test assessing children's narrative comprehension skills were expected to be too difficult for the younger children (aged, e.g., 4–5 years). In other words, at the time the tests were developed we anticipated that not every test would be to an equal extent appropriate for all deaf children in primary education. As it was hard to predict to what extent a particular test was too difficult or too easy for deaf children in a particular age-group, all tests were administered to the whole group of deaf children. It is important to note that all the tests in which the expressive skills of children were assessed were scored "live" by the test administrator. This choice was explicitly made to make future administration of the tests by SLN teachers at schools for the deaf after the completion of the project as less time consuming as possible. As re-watching the children's responses on the expressive tests from videotape after the test administration would considerably increase the time SLN teachers need to administer and score the sign language tests, we decided that the children's responses had to be scored during the administration of the tests.

In the next part of this article, each sign language test will be described. Before the language tests were developed, an administrative shell was programmed in Macromedia Director to make the administration of the SLN tests easier. Before assessing the SLN tests, the administrator typed the child's name, gender, and age. This information was stored in a database and was updated after every test administration, as the results of the receptive tests were also stored in this database.

The SLN Subtests

In the receptive phonology task, two signs were consecutively presented on a computer screen. Children were instructed to decide whether the signs had the same meaning (press the green button with the mouse) or a different meaning (press the red button). The two signs were produced by two deaf female native signers, one signer always producing the first sign and one signer always producing the second sign. A proportion of the signs were minimal pairs in SLN and different only with respect to one phonological parameter: the spoken component, the handshape, or the movement. The rationale behind the task was that if children had not yet acquired this phonological parameter in SLN, they would have difficulties in discriminating these minimal pairs. The receptive phonology task consisted of 36 items.

The children's expressive phonological skills were assessed in an imitation task. In this task, a sign was presented on a computer screen. Children were instructed to repeat the sign. The test administrator judged the correctness of one phonological parameter of the sign produced by the child (e.g., handshape, movement, oral component). As we pointed out earlier, the expressive tests were scored during the test administration itself. This choice was made to make the test administration less time consuming, especially after the completion of the project. As we expected that test administrators could only effectively judge one phonological parameter per item, we decided to select one phonological parameter for each of the signs. The test materials also included an information sheet for the test administrator on which the possible correct responses (phonetic variations) for each sign were depicted. These information answer sheets were developed in collaboration with SLN linguistics and native signers.¹ Figure 1 shows the example from this information

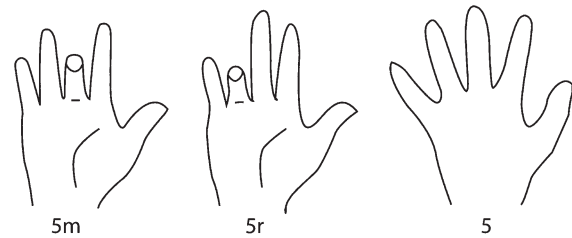


Figure 1 An example of a test item on the administrative sheet of the expressive phonology test.

sheet for the SLN sign 'RING' (Ring), which can be produced with the '5m', '5r' or '5' handshape.

In the receptive vocabulary task, a sign was presented on a screen followed by four pictures (see Figure 2). Children were instructed to select the picture that matched the meaning of the sign by selecting the picture with the mouse button. In all, the test consisted of 61 items. One of the major problems in developing a sign vocabulary test concerns the iconicity of signs (Jansma et al., 1997; White & Tischler, 1999). The problem is that children who encounter a sign that they have not acquired yet may exploit the iconic features of the sign to correctly guess its meaning and select the appropriate picture. We basically used two strategies in an attempt to reduce this problem: (a) we added foil pictures (that did not match the sign in meaning) that resembled the shape of the sign and (2) the picture that matched the sign in meaning was drawn from such a perspective that its shapes no longer (or as little as possible) resembled the iconic features of the sign.

To assess to what extent we had successfully managed to reduce unwarranted effects of iconicity from the test, the receptive vocabulary test was administered to a group of 28 hearing children, aged 11–12, with no knowledge of SLN.² The percentage of correct responses for these children ranged from 21.3% to 42.6%, with an average of 33.5%. This percentage differed significantly from chance (25%), as indicated by a *t* test, $t(27) = 7.87, p < .001$. This finding may suggest that hearing children can still effectively exploit the iconicity of signs. However, it is important to note that hearing children without the knowledge of SLN may not only have exploited the iconicity of signs but its spoken component as well. The SLN test consisted predominantly of nouns that are usually



Figure 2 An example of a test item of the receptive vocabulary test.

accompanied by a spoken component in SLN that, in citation form, often consists of the whole word. After the experiment, some of the hearing children reported that they used the spoken component to guess the correct picture. In other words, it is possible that the group of hearing children without SLN knowledge may have scored above chance because they could exploit the spoken component instead of the iconic features of the sign. Nevertheless, even though the hearing children performed above chance, the problem was much less pronounced as in previous studies (Jansma et al., 1997; White & Tischler, 1999).

The expressive vocabulary skills were assessed in two tasks: the expressive vocabulary I and the expressive vocabulary II tasks. In the expressive vocabulary I task, a picture was presented on a screen. Children were instructed to name the picture in SLN. The test consisted of 54 items. The test administrator scored whether the children had produced the correct response in SLN. In the expressive vocabulary II task, a sign was presented in SLN. Children were instructed to describe its meaning. The test consisted of 40 items. Again, the test administrator wrote down whether the deaf child had successfully managed to describe the meaning of the sign in SLN during the administration.

The children's receptive and expressive morphosyntactic skills were assessed in two tasks. In both tasks, a variety of morphosyntactic rules of SLN were tested (e.g., verb agreement, modifications of verbs for aspect, classifier verbs of motion and location). In the receptive morphosyntactic task, a phrase or sentence in SLN was presented on the screen, followed by four pictures. Children were instructed to select the picture that matched the phrase or sentence.

In the expressive morphosyntax task, a picture was presented on a screen (see Figure 3). Then, an SLN video appeared next to the picture, and the picture's content was described in SLN. Finally, another picture was presented on the screen, and children were instructed to describe the picture in SLN. The test administrator scored whether the child had successfully described its content. Again, the test materials also included an information sheet for test administrator on which the possible correct responses for each picture were described.³ The first picture and its description in SLN were used to elicit an appropriate response. The expressive morphosyntax task consisted of 24 items.

The narrative comprehension and production skills of the children were assessed in two tasks. In the narrative comprehension task, five SLN stories were presented on the computer screen. After each story, four questions were presented in SLN on the screen, and children were instructed to answer these questions. Some of these questions referred to information literally mentioned in the stories. Other questions were either questions requiring gap-filling or text-connecting inferences (Cain & Oakhill, 1999).⁴ The test administrator scored whether the children

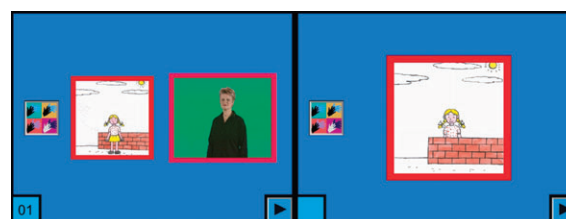


Figure 3 An example of a test item of the expressive morphosyntax test.

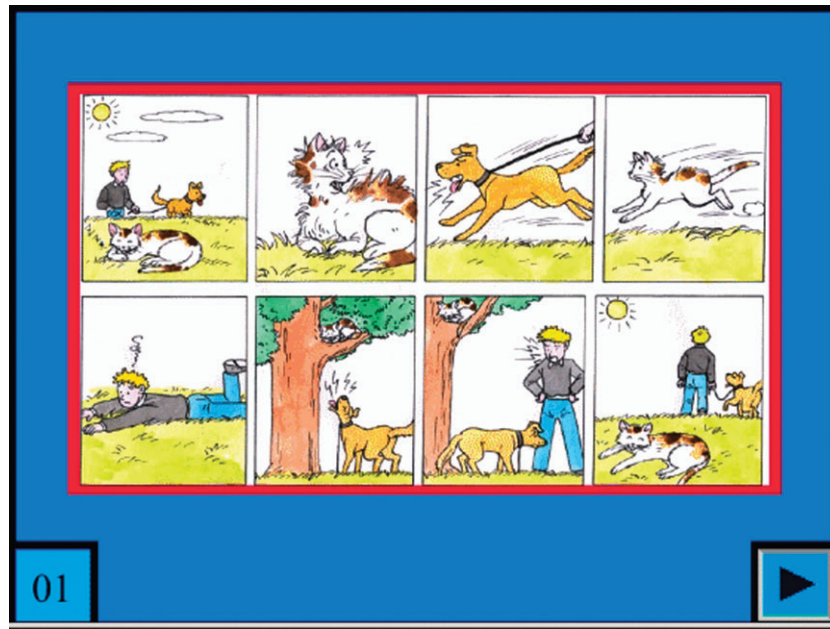


Figure 4 A story from the narrative production task.

had correctly answered each question during the administration. The narrative comprehension task consisted of five stories and 20 questions. The average length of the stories was 53 s (range 39–83 s).

In the narrative production task, a sequence of eight static cartoon illustrations was depicted on the screen (see Figure 4). Children were instructed to study the story. Then, the depicted story disappeared from the screen, and children were asked to retell the story in SLN.

The Norming Study

The aim of the norming study was to collect data on the basis of which norms could be developed for each age-group (Group 1: ages 4.0–4.11, Group 2: ages 5.0–5.11, etc.). A classical issue in the development of assessment instruments for children is a decision on who constitutes the norm (Singleton & Supalla, 2003). Ideally, separate norms are developed for subgroups of the total sample when it is known that the developmental path of the acquisition of the language in question is different for this subgroup. To illustrate, spoken language tests sometimes have separate norm for boys and girls (Zink & Lejaegere, 2002), or for children who are born in bilingual families or in fam-

ilies in which the parents do not speak the language in question (Verhoeven & Vermeer, 2001). In our case, deaf children of deaf parents would constitute such a subgroup as deaf children of deaf parents, on average, tend to have a higher level of proficiency in comparison with deaf children of hearing parents (Boudreault & Mayberry, 2006; Hoiting, 2005; Herman & Roy, 2006; Maller et al., 1999). Johnson (2004), very rightly, has questioned the appropriateness of using an entire population of children who are exposed to sign language to develop normative scores. With the exception of the sign language assessment instrument developed by Anderson and Reilly (2002), which was normed with exclusively deaf children of deaf parents, all the norm-referenced assessment instruments that have now been developed have used combined groups of native (deaf and hearing) children and nonnative deaf children to develop normative scores (Herman et al., 1999; Maller et al., 1999). At present, our group of deaf children is unfortunately not large enough to develop separate norms for subgroups, and it is unlikely that norms for every subgroup of deaf children (especially deaf children of deaf parent) can ever be developed within our country. The norms were extracted on the basis of all the deaf children who participated in the norming study.

Table 2 Characteristics of the children who participated in the norming study

Ethnicity of parents	Parental hearing status	Boys	Girls	Total
Dutch	Deaf/hard of hearing	12	13	25
	Hearing	108	93	201
Other	Deaf/hard of hearing	5	2	7
	Hearing	49	48	97
		174	156	330

Participants

In a 3-year norming study, the nine tests were administered to a group of deaf children, aged 4–12 years, from seven schools (out of eight) for deaf children in the Netherlands. The schools all provide bilingual education for deaf children. All children had a hearing loss more than 80 dB on the best ear (unaided), had normal nonverbal intelligence, and did not have additional known handicaps.⁵

Table 2 shows some of the characteristics of the children who participated to the norming study. A total of 330 children were tested. One hundred sixty-three children were tested in 3 consecutive years; 76 children were tested twice in 2 consecutive years, whereas 91 children were tested once.

Procedure

The tests were administered by deaf/hard-of-hearing third- and fourth-year students of the Institute SLN where they were trained to become sign language interpreters or sign language teachers. Prior to their participation, they were trained in the administration of the tests by the first author. Each year, a training day was organized at the Institute SLN, and it consisted of two parts. In the morning, the general aim of sign language assessment was explained, and every test was shown and explained to the (future) administrators. In the afternoon, several training sessions were conducted with the expressive tests. In turn, the students administered the expressive tests to fellow students in time slots of approximately 30 min. Subsequently, questions on the administration of the tests were discussed groupwise. The test administrators were paid for their participation in the project. The tests were administered individually.

Constructing the Norms

The goal of the norming study was to define five categories for each age-group, following the TAK (“Taaltest Alle Kinderen” [language test for all children]; Verhoeven & Vermeer, 2001). On the basis of the raw scores of all the children in each age-group, the following five intervals were defined: (A) good to excellent (above the 75th percentile), (B) average to good (between the 50th and 75th percentile), (C) moderate to average (between the 25th and 50th percentile), (D) poor to moderate (between the 10th and 25th percentile), and (E) very poor to poor (below the 10th percentile). Table 3 shows the norms for the 10-year-old (10.0–10.11) children in relation to their raw test scores.

Psychometric Properties of the Instrument

Reliability

Reliability refers to the stability of measurements of the tests (Law, Boyle, Harris, Harkness, & Nye, 1998). The question really is to what extent the circumstances under which the assessment has taken place (different conditions, different test administrators) will affect the measurement. For the SLN tests, three aspects of the reliability of the sign language tests were assessed: the internal consistency of the tests, the test–retest reliability, and the interrater reliability.

Internal consistency. Cronbach’s alphas were computed for each age-group and each test separately (see Table 4). Across all age-groups and tests, there was variance on the Cronbach’s alpha values, ranging from .60 to .96. The average alphas for each test were very good ($>.90$) for the receptive vocabulary task and the receptive morphosyntax task; good ($>.80$) for the receptive phonology task, the expressive vocabulary I task, expressive vocabulary II task, the expressive morphosyntax task, and the narrative comprehension task; and acceptable ($>.70$) for the expressive phonology task and the narrative production task.

Test–retest reliability. A second method to assess the reliability of the tests is to investigate test–retest reliability. Usually, tests are readministered within

Table 3 Examples of the extracted norms

Test	E	D	C	B	A
Receptive phonology	0–28	29–30	31–33	34	35–36
Expressive phonology	0–25	26–27	28–30	31–32	33–34
Receptive vocabulary	0–22	23–42	43–45	46–49	50–61
Expressive vocabulary I	0–25	26–30	31–35	36–41	42–54
Expressive vocabulary II	0–8	0–10	11–13	14–20	21–40
Receptive morphosyntax	0–24	25–29	30–34	35–38	39–47
Expressive morphosyntax	0–9	10–13	14–18	19–22	23–24
Narrative comprehension	1–7	8–12	13–15	16–18	19–20
Narrative production	0–12	13–15	16–18	19	20

a relatively short time interval after the first assessment, for instance, within 3 or 6 months. Such data were not available for the sign language tests. However, a large proportion of the children were tested in 2 or 3 consecutive years. We have used those data to estimate the test–retest reliability. The correlations (Spearman's rho) of children's scores on two consecutive test administrations were .53, .56, .71, .80, .77, .73, .74, .83, and .81, respectively, for the tests assessing receptive phonology, expressive phonology, receptive vocabulary, expressive vocabulary I, expressive vocabulary II, receptive morphosyntax, expressive morphosyntax, narrative comprehension, and narrative production (all $ps < .001$). Especially for the receptive and expressive phonology tests, test–retest reliability was only moderate. Note that using such a large time interval (1 year) will probably have negatively affected the test–retest reliability as there will be differences between children's acquisition of SLN skills between the test administrations. In other words, this procedure has presumably resulted in an underestimation of the test–retest reliability.

Interrater reliability. The expressive tasks were scored by the test administrator during test adminis-

tration itself. The expressive tasks were also videotaped to investigate the interrater reliability of the scoring. For each of the five expressive tasks, another group of 13 test administrators scored a randomly selected group of children within a particular age-group for the second time but now from videotape.

Table 5 lists the number of children whose test performance was assessed twice for each of the five expressive tests, and the correlation between the scores of the first and second group of administrators. As shown in Table 5, the correlations (Spearman's rho) ranged from .78 (narrative production) to .92 (expressive vocabulary II), which can be called high.

Validity

The validity of a test refers to the question to what extent a test measures what it intends to measure. The concurrent validity of the sign language tests could not be determined because there are no assessment instruments of SLN available, like in most other countries (Haug & Mann, 2008). However, the construct validity and predictive validity of the test could be examined.

Table 4 Cronbach's alpha values for each of the tests and age-groups

Test	4	5	6	7	8	9	10	10+	Average
Receptive phonology	.90	.87	.85	.85	.89	.81	.75	.71	.83
Expressive phonology	.75	.79	.81	.73	.60	.69	.71	.70	.72
Receptive vocabulary	.90	.94	.96	.95	.96	.96	.95	.96	.95
Expressive vocabulary I	.87	.88	.88	.90	.90	.87	.86	.85	.88
Expressive vocabulary II	.82	.74	.78	.78	.87	.90	.91	.92	.84
Receptive morphosyntax	.90	.92	.94	.93	.94	.92	.88	.92	.92
Expressive morphosyntax	.86	.85	.89	.90	.88	.78	.86	.84	.86
Narrative comprehension	.87	.81	.87	.90	.91	.86	.86	.88	.87
Narrative production	.82	.86	.90	.90	.76	.68	.72	.68	.79

Table 5 Interrater reliability (as estimated by Spearman's rho) of the expressive Sign Language of the Netherlands tests

Test	Number	Age-group	Correlation
Expressive phonology	25	5.0–5.11	.87 (.000)
Expressive vocabulary I	19	6.0–6.11	.91 (.000)
Expressive vocabulary II	19	7.0–7.11	.92 (.000)
Expressive morphosyntax	22	8.0–8.11	.86 (.000)
Narrative production	19	9.0–9.11	.78 (.000)

A way to assess construct validity is to investigate how characteristics of the participants that are known to affect the children's acquisition of (sign) language also affected the children's scores on the SLN tests (e.g., Maller et al., 1999). Three characteristics of the children were investigated: their age, their gender, and the hearing status of their parents.

One minimal requirement for any developmental test assessing children's language proficiency is that there is a significant relation between children's test scores and their age (van Eldik et al., 2003). For each sign language test, the correlation between the children's ages (in months) and their test scores was computed. The correlations (Spearman's rho) between the children's age and their test scores were, respectively, .61 for the receptive phonology task, .60 for the expressive phonology task, .73 for the receptive vocabulary task, .75 for the expressive vocabulary I task, .73 for the expressive vocabulary II task, .70 for the receptive morphosyntax task, .74 for the expressive morphosyntax task, .76 for the narrative comprehension task, and .69 for the narrative production task (all $ps < .001$).

The second construct validity variable that was investigated was the children's gender. Generally, females outperform males on verbal tasks, especially in early childhood (Maccoby & Jacklin, 1974; McCarthy, 1954; Verhulst-Schlichting, Morelli-Kaiser, & Peddemors-Boon, 1987; Zink & Lejaegere, 2002). For deaf children, however, this gender effect has not been consistently reported in the literature yet. To illustrate, Maller et al. (1999) found no gender effect in a sample of 80 deaf children, aged 6–12, on the test scores of the American Sign Language Proficiency Assessment. In contrast, Herman and Roy (2006) observed a gender effect in an analysis of the

Table 6 Parental hearing status and gender effects in the Sign Language of the Netherlands tests

Test	Parental hearing status	Gender
Receptive phonology	.41**	.21**
Expressive phonology	.34**	.31***
Receptive vocabulary	.50***	.37***
Expressive vocabulary I	.62***	.21**
Expressive vocabulary II	.83***	.22**
Receptive morphosyntax	.61***	.27***
Expressive morphosyntax	.76***	.19*
Narrative comprehension	.68***	.45***
Narrative production	.35*	.20*

* $p < .05$, ** $p < .01$, *** $p < .001$.

BSL test scores of 163 girls and 153 boys. In the Herman and Roy study, girls were also found to consistently outperform boys. In the present study, there were 156 girls and 174 boys. As there were differences between the average age of girls and boys, the test scores of each child were transformed into z scores based upon the mean and standard deviation of the age-group of the children. Table 6 lists the differences in the z scores of the girls and boys who participated in the norming study (positive values indicate that girls outperformed boys), revealing that girls outperformed boys on every SLN test. As shown in Table 6, analyses of variance revealed that the group of girls outperformed the group of boys on each of the nine SLN tests.

The third and final variable that was investigated as an indicator of construct validity of the SLN test was the hearing status of the parents of the children. Deaf children with deaf parents often have better signing skills than deaf children with hearing parents (Boudreault & Mayberry, 2006; Hoiting, 2005; Herman & Roy, 2006; Maller et al., 1999). We therefore analyzed to what extent the deaf children of deaf parents in our study also had higher test scores on the SLN tests. In the present study, 32 children had one or two deaf parents, whereas 298 deaf children had hearing parents. As there were differences between the average age of the deaf children of deaf parents and the average age of the deaf children of hearing parents, the test scores of each child were transformed into z scores based upon the mean and standard deviation of the age-group of the children. Table 6 lists

the differences in the z scores of deaf children with deaf parents and deaf children with hearing parents in the present sample (positive values indicate that deaf children of deaf parents outperformed deaf children of hearing parents). As shown in Table 6, analyses of variance revealed that the group of deaf children of deaf parents significantly outperformed the group of deaf children of hearing parents on each of the nine SLN tests.

Finally, the predictive validity of the sign language assessment instrument was explored. The newly developed SLN test has also been used in various research projects in which the predictive relation between sign language proficiency and written/spoken language skills in Dutch was studied (Ormel, 2008). In a longitudinal study, Ormel found that the sign language vocabulary scores derived from the receptive vocabulary test we developed for a group of 62 deaf children from bilingual education programs were significantly correlated with the reading comprehension scores 1 and 2 years later, replicating and extending the results obtained in previous studies (Chamberlain & Mayberry, 2000, 2008; Dubuisson, Parisot, & Vercaigne-Menard, 2008; Hoffmeister, 2000; Mann, 2007; Niederberger, 2008; Padden & Ramsey, 2000; Strong & Prinz, 1997, 2000). Again, obtaining the same results as in previous studies can be taken as evidence that the predictive validity of the sign language assessment instrument is high.

In other words, although there are little means available to investigate the concurrent validity of the sign language tests, the construct and predictive validity of the tests could be investigated. Three variables that are known to affect children's sign language development, age, gender, and parental hearing status, also affected the children's test scores in the expected direction.⁶

Discussion

Herman (1998) conducted a survey in the United Kingdom in schools that used BSL in their educational program for deaf children. Most of the respondents indicated that there was clear need for standardized assessment instruments at their schools. Respondents to a survey conducted by Mann and

Prinz (2006) in the United States expressed the same need for standardized assessment instruments. In the Netherlands, schools for the deaf that use SLN in their educational program have also struggled with the assessment of deaf children's sign language skills. The work we report upon in this article was initiated to fulfill this need.

Nine computer-based sign language tests were developed and administered to a group of 330 deaf children, aged 4–12, in a 3-year norming study. The data collected in the norming study were used to develop norms for each age-group. We also examined some of the psychometric properties of the SLN tests. The analyses of the reliability of the test revealed that the average internal consistency of the assessment tests was "acceptable" to "good." In general, the average internal consistency of the receptive tasks (.89) was higher than that of the expressive tasks (.82). The expressive tasks required the test administrator to evaluate the correctness of the children's responses. This will always lead to an increase in the standard error of measurement. Although interrater reliability was sufficiently high (average .87), receptive language tests will generally be more reliable than productive language tests.

There were little means available to evaluate the concurrent validity of the SLN tests. We are currently involved in a small-scale project in which the concurrent validity of the SLN tests is investigated by comparing children's scores on the SLN tests with the scores on the "NGT-OP" (NGT Observatiepakket voor Peuters [SLN observation list for toddlers]; Baker & Jansma, 2005), an assessment instrument for toddlers and infants. The construct validity of the tests was examined by investigating the relationship between the children's scores on the SLN tests and the children's age, gender, and parental hearing status. These investigations revealed that all three variables affected the children's scores on every subtest in the expected direction: older children outperformed younger children, deaf children of deaf parents outperformed deaf children of hearing parents, and girls outperformed boys. These psychometric investigations demonstrate that the construct validity of the SLN tests is high. At the same time, they have also revealed the necessity to construct separate norms as

Table 7 Self-evaluation on the basis of Haug's (2005) criteria

Criterion	Self-evaluation
1. Robust psychometric properties	Variable
2. Clarity on specificity of tests	Yes
3. Choice of features supported by linguistic research	No
4. Language samples elicited in various contexts	No
5. Design permitting efficient administration/ analyses	Yes
6. Use in educational settings	Yes
7. Use for research purposes	Yes
8. Assessment of receptive and expressive skills	Yes
9. Development with assistance of deaf researchers	Yes
10. Applicability to adults and children	No
11. Applicability to broad age range	Yes
12. Assessment of language proficiency and communicative competence	No
13. General availability	Yes
14. Adaptation to issues such as iconicity	Yes
15. Availability of age norms	Yes

functions of children's gender and parental hearing status, as pointed out by Johnson (2004). The data obtained in the present project do not allow the construction of norms for these separate subgroups, and the bottom line is that for deaf children of deaf parents, the development of separate norms is not feasible in our country. With an eye on predictive validity, it was shown that the sign language skills of deaf children predicted their literacy development throughout the primary grades.

Haug (2005) has proposed a list of criteria to evaluate the strengths and weaknesses of sign language assessment instruments. The criteria and our self-evaluation for each of the criterion for the SLN tests with respect to these criteria are listed in Table 7.

There are four criteria to which our self-evaluation of the assessment instrument leads to a negative answer: Criteria 3, 4, 10, and 12. Of those criteria, we consider Criterion 3 (choice of features supported by linguistic research) as the most threatening one. There are 10 criteria on which our self-evaluation leads to positive answer: Criteria 2, 5, 6, 7, 8, 9, 11, 13, 14, and 15. For some of these criteria, additional remarks from our side seem to be necessary. Because the tests were

developed and normed, several schools in the Netherlands have started to use the assessment instrument in their school. Although the design permits efficient administration and analyses (Criterion 5) in comparison to many tests assessing spoken or written language skills, the bottom line is that administering the assessment instrument takes approximately 2 hr for each child and requires knowledge (for the expressive tests) of the phonology and grammar of SLN. There is a shortage of people in the schools in the Netherlands who are qualified to administer the language tests and those people who are qualified (e.g., SLN teachers, SLN consultants) usually do not have the time available that is required for a yearly assessment of the expressive and receptive skills of all deaf children in their school. In other words, in practice, incorporating the sign languages assessment instrument in the schools is not easy.

Furthermore, even though norms are now available for the assessment instrument (Criterion 15), these norms do not take into account the parental status of the children and their gender, variables that were found to affect the signing skills of the children in the present study. In addition, some of the variables that are likely to affect the children's signing skills (e.g., the social economic status of their parents) were not taken into account in the present study. This implies that the test results should be interpreted carefully. In the manual of the SLN tests, we have added several hypothetical case studies to clarify to test administrators at schools for the deaf how to interpret the test scores of the children.

Finally, for one of the criteria (Criterion 1: robust psychometric properties), our self-evaluation was not unequivocal. Not every psychometric property could be studied (concurrent and predictive validity) and the psychometric properties that were investigated varied between "moderate" and good. Nevertheless, given the huge challenge that we have faced in the development of the assessment instrument, the results seem quite satisfactory.

Educational Perspective

Because the development of the sign language tests was completed (Hermans, Knoors, & Verhoeven,

2007), most of the schools for deaf children now use the SLN tests to assess children's sign language skills. There are, however, differences in the approaches schools have taken to incorporate the sign language tests to assess their pupils' sign language skills. To illustrate, one of the schools has opted for yearly administration of the SLN tests to all deaf children. Although this is clearly a huge investment, this is the only guarantee that children who fall behind in the acquisition of SLN can easily be detected. Other schools have chosen to administer the language tests only when they suspect that a child's SLN development is delayed or hampered in comparison with his/her deaf peers. Yet, other schools have chosen an approach in between, for instance, yearly administration of some of the SLN tests to all deaf children in the school. These data are currently being collected in an attempt to strengthen the normative scores and study the psychometric properties of the SLN instrument in further depth.

Notes

1. The administrative sheet listed the acceptable phonetic variations (for the phonological parameter in question). This administrative sheet was developed in collaboration with Crasborn, van der Kooij, and Emmerik of the Radboud University. For each of the signs, several deaf informants indicated how a particular sign could and could not be realized in SLN. Furthermore, for some, but not all of the signs, database information was available on the different phonetic variations of the sign. This information was used to construct the administrative sheet.

2. As pointed out by one of the reviewers, administering the sign vocabulary test to a group of hearing children is difficult as this group of participants may use a very different strategy to conduct this particular task.

3. The administrative sheet was developed in collaboration with deaf and hearing coworkers of the Institute NGT in Utrecht. Two discussion sessions were organized in Utrecht, in which it was discussed how the 24 items in the expressive morphosyntactic task could and could not be expressed in SLN. This information was used to construct the administrative sheet.

4. In gap-filling inferences, the child is required to integrate his/her own knowledge with information in the narrative to fill in details not explicitly stated. In text-connecting inferences, children must integrate information explicitly mentioned in the narrative to link ideas in two sentences.

5. For selection purposes, this information was obtained from the pupil's files in the schools for the deaf, and often consisted of test results from the 'Snijders-Oomen Niet-Verbale Intelligentie tests - Revisie' (Snijders-Oomen Nonverbal Intelligence tests - Revised) by Snijders, Tellegen and Laros (1998).

6. In addition, we administered Raven's Coloured Progressive Matrices (CPM) test to the children who participated in the norming study to investigate the relationship between the children's test scores and their nonverbal intelligence (as indicated by their score on the Raven CPM). The correlations between the children's test scores on the nine SLN tests and the Raven CPM varied between .03 and .34, with an average of .20, suggesting that cognitive skills like nonverbal intelligence did not strongly determine the children's test scores in the SLN tests.

References

- Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, 7, 83–106.
- Baker, A. E., & Jansma, S. (2005). *NGT-OP: Een screening instrument voor peuters in de Nederlands Gebaren taal en handleiding* [NGT-OP: A screening instrument for young children and handbook]. Amsterdam: University of Amsterdam.
- Bos, H. F. (1990). Person and location marking in SLN: Some implications of a spatially expressed syntactic system. In S. Prillwitz & T. Vollhaber (Eds.), *Current trends in European sign language research* (pp. 231–246). Hamburg, Germany: Signum.
- Bos, H. F. (1993). Agreement and prodrop in Sign Language of the Netherlands. In F. Drykoningen & K. Hengeveld (Eds.), *Linguistics in the Netherlands* (pp. 37–48). Amsterdam/Philadelphia: Benjamins.
- Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, 21, 608–635.
- Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure. *Reading and Writing*, 11, 489–503.
- Chamberlain, C., & Mayberry, R. I. (2000). Theorizing about the relation between American Sign Language and Reading. In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 221–259). Hillsdale, NJ: Erlbaum.
- Chamberlain, C., & Mayberry, R. I. (2008). American Sign Language syntactic and narrative comprehension in skilled and less skilled readers: Bilingual and bimodal evidence for the linguistic basis of reading. *Applied Psycholinguistics*, 29, 367–388.
- Coerts, J. (1992). *Nonmanual grammatical markers: An analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands*. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam.
- Crasborn, O. (2001). *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. Unpublished doctoral dissertation, LOT, Utrecht.
- Dubuisson, C., Parisot, A.-M., & Vercaigne-Menard, A. (2008). Bilingualism and deafness: Correlations between deaf students' ability to use space in Quebec Sign Language and their reading comprehension in French. In C. Plaza Pust & E. Moralez-Lopez (Eds.), *Sign bilingualism: Language development, interaction, and maintenance in sign language contact situations* (pp. 51–71). Amsterdam/Philadelphia: Benjamins.

- Harder, R., & Schermer, T. (1986). A first phonological analysis of handshapes in the Sign Language of the Netherlands (SLN). In B. Tervoort (Ed.), *Signs of life. Proceedings of the 2nd European congress on sign language research* (pp. 47–51). Amsterdam: Dutch Foundation for the Deaf and Hearing-Impaired Child.
- Haug, T. (2005). Review of sign language assessment instruments. *Sign Language & Linguistics*, 8, 61–98.
- Haug, T., & Mann, W. (2005). Projekt Gebardensprachtest—Entwicklung von Testverfahren zur Deutschen Gebardensprache für gehörlose Kinder und Jugendliche [Project sign language test—Development of test instruments for German sign language for deaf children and adolescent]. *Das Zeichen*, 71, 370–380.
- Haug, T., & Mann, W. (2008). Adapting tests of sign language assessment for other sign languages—A review of linguistic, cultural, and psychometric problems. *Journal of Deaf Studies and Deaf Education*, 13, 138–147.
- Herman, R. (1998). The need for an assessment of deaf children's signing skills. *Deafness & Education*, 4, 3–8.
- Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL development—Receptive skills test*. Coleford, England: Forest Bookshop.
- Herman, R., & Roy, P. (2006). Evidence from the wider use of the BSL receptive skills test. *Deafness and Education International*, 8, 33–47.
- Hermans, D., Knoors, H., & Verhoeven, L. (2007). *An assessment instrument for Sign Language of the Netherlands*. Viat: Sint-Michielsgestel.
- Hoffmeister, R. J. (1999). *American Sign Language assessment instrument (ASLAI)*. Unpublished test, Center for the Study of Communication & the Deaf, Boston University, Boston.
- Hoffmeister, R. J. (2000). A piece of the puzzle: ASL and reading comprehension in deaf children. In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 143–163). Hillsdale, NJ: Erlbaum.
- Hoiting, N. (2005). Deaf children are verb attenders: Early sign vocabulary development in Dutch toddlers. In B. Schick, M. Marschark, & P. Spencer (Eds.), *Advances in sign language development of deaf children* (pp. 161–188). New York: Oxford University Press.
- Jansma, S., Knoors, H., & Baker, A. E. (1997). Sign language assessment: A Dutch project. *Deafness & Education*, 21, 39–46.
- Johnson, T. (2004). The assessment and achievement of proficiency in a native sign language within a sign bilingual program: The pilot Auslan receptive skills test. *Deafness and Education International*, 6, 57–81.
- Knoors, H. (1992). *Exploratie van de gebarenruimte*. Unpublished doctoral Dissertation, Eburon, Koninklijke Ammanstichting Rotterdam, Delft.
- Knoors, H. (2007). Educational responses to varying objectives of parents of deaf children: A Dutch perspective. *Journal of Deaf Studies and Deaf Education*, 12, 243–232.
- Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: A systematic review of the literature. *Health Technology Assessment*, 2(9), 1–184.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Maller, S., Singleton, J., Suppalla, S., & Wix, T. (1999). The development and psychometric properties of the American Sign Language Proficiency Assessment (ASL-PA). *Journal of Deaf Studies and Deaf Education*, 4, 249–269.
- Mann, W. (2007). German deaf children's understanding of referential distinction in written German and German Sign Language. *Educational & Child Psychology*, 24, 59–76.
- Mann, W., & Prinz, P. (2006). The perception of sign language assessment by professionals in deaf education. *American Annals of the Deaf*, 151, 356–370.
- McCarthy, D. (1954). Language development in children. In L. Carmichael (Ed.), *Manual of child Psychology*. New York: Wiley.
- Mountry, J. (1993). *Signed language development checklist—Training manual*. Princeton, NJ: Educational Testing Service.
- Mountry, J. (1994). *Signed language development checklist*. Princeton, NJ: Educational Testing Service.
- Niederberger, N. (2008). Does the knowledge of a natural sign language facilitate Deaf children's learning to read and write? Insights from French Sign Language and written French data. In C. Plaza Pust & E. Moralez-Lopez (Eds.), *Sign bilingualism: Language development, interaction, and maintenance in sign language contact situations* (pp. 39–50). Amsterdam/Philadelphia: Benjamins.
- Ormel, E. (2008). *Visual word recognition in bilingual deaf children*. Unpublished doctoral dissertation, University of Nijmegen, Nijmegen.
- Padden, C., & Ramsey, C. (2000). American Sign Language and reading ability in deaf children. In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 165–189). Hillsdale, NJ: Erlbaum.
- Prinz, P., Strong, M., & Kuntze, M. (1994). *The test of ASL*. Unpublished test, San Francisco State University, California Research Institute, San Francisco.
- Schembri, A., Wigglesworth, G., Johnston, T., Leigh, G., Adam, R., & Barker, R. (2002). Issues in development of the test battery for Australian sign language morphology and syntax. *Journal of Deaf Studies and Deaf Education*, 7, 18–40.
- Singleton, J. L., & Supalla, S. (2003). Assessing children's proficiency in natural signed languages. In M. Marschark & P. Spencer (Eds.), *Oxford handbook of deaf studies, language and education* (pp. 289–302). New York: Oxford University Press.
- Schermer, G. M. (1983). Het communicatiesysteem van volwassen doven in Nederland. In B. Th. Tervoort (Ed.), *Hand over hand* (pp. 9–24). Muiderberg, The Netherlands: Coutinho.
- Schermer, G. M. (1990). *In search of a language. Influences from spoken Dutch on Sign Language of the Netherlands*. Unpublished doctoral dissertation, Universiteit van Amsterdam, Amsterdam.
- Schermer, G. M., Fortgens, C., Harder, R., & de Nobel, E. (1991). *De Nederlandse Gebarentaal*. Twello, The Netherlands: Van Tricht.
- Schermer, T., & Koolhof, C. (1990). The reality of time-lines: Aspects of tense in Sign Language of the Netherlands. In S. Prillwitz & T. Vollhaber (Eds.), *Current trends in European sign language research* (pp. 295–305). Hamburg, Germany: Signum.

- Snijders, J. Th., Tellegen, P. J., & Laros, J. A. (1998). *Snijders-Oomen Niet-verbale Intelligentietest* [Snijders-Oomen test of nonverbal intelligence]. Amsterdam: Boom.
- Strong, M., & Prinz, P. M. (1997). A study of the relationship between ASL and English literacy. *Journal of Deaf Studies and Deaf Education*, 2, 37–46.
- Strong, M., & Prinz, P. M. (2000). Is American Sign Language related to English Literacy? In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 131–141). Hillsdale, NJ: Erlbaum.
- van der Kooij, E. (2002). *Reducing phonological categories in Sign language of the Netherlands. Phonetic implementation and iconic motivation*. Unpublished doctoral dissertation, LOT, Utrecht.
- van Eldik, M. C. M., Schlichting, J. E. P. T., Lutje Spelberg, H. C., van der Meulen, B. F., & Van der Meulen, S. J. (2003). *Reynell test voor Taalbegrip* [Reynell test for language comprehension]. Lisse, The Netherlands: Swets.
- van Gijn, I. (2004). *The quest for syntactic dependency. Sentential complementation in Sign Language of the Netherlands*. Unpublished doctoral dissertation, LOT, Utrecht.
- Verhoeven, L., & Vermeer, A. (2001). *Taaltoets alle kinderen* [Language test for all children]. Arnhem, The Netherlands: Citogroep.
- Verhulst-Schlichting, J. E. P. T., Morelli-Kaiser, M., & Peddemors-Boon, M. (1987). Sex and family background in early language acquisition. In D. Brouwer & D. de Haan (Eds.), *Women's language, socialisation and self-image* (pp. 24–65). Dordrecht, The Netherlands: Foris.
- White, A., & Tischler, S. (1999). Receptive sign vocabulary tests: Tests of sign-word vocabulary or iconicity? *American Annals for the Deaf*, 144, 334–338.
- Zink, I., & Lejaegere, M. (2002). *N-CDIs: lijsten voor communicatieve ontwikkeling* [N-CDIs: Lists for communicative development]. Leuven, The Netherlands: Acco.
- Zwitserslood, I. (2003). *Classifying hand configurations in Nederlandse Gebarentaal*. Unpublished doctoral dissertation, LOT, Utrecht.

Received May 5, 2009; revisions received September 22, 2009; accepted October 2, 2009.