The language-screening instrument SNEL

Luinge, Margreet Roelien

Chapter 6

The language-screening instrument SNEL

Abstract

The aim of this study was to design a valid and accurate language-screening instrument (SNEL) for identifying possible language problems in children from one to six years of age, which is intended to improve the referral of at-risk children for further diagnosis.

A scale of several language milestones was constructed according to the nonparametric Mokken Item Response Model for dichotomous responses. The external validation of the scaled language milestones of the SNEL scale was examined by means of sample independence, the relationship between age and scaled language milestones, the relationship between age and SNEL scores, and the relationship between SNEL and the gold standard. The ability of SNEL to identify children with language problems in both the normal population and in a clinical population was also examined.

The scaling results showed that the scalability of the SNEL-scale was strong (H = 0.95) and its reliability was high (Rho = 0.96). The external validation showed that the ordering of milestones was the same for both boys and girls and that it did not differ across either age groups or geographical regions in the Netherlands. The results showed further that the scaled language milestones as well as the SNEL scores increased with age, and that SNEL measured the same construct, language production, that was measured by the reference test. Accuracy measured against parental reports of the language development of children between the ages of twelve to seventy-two months was satisfactory. SNEL proved sufficiently sensitive to detect children with possible language problems. In conclusion, the unidimensional and sample-independent SNEL scale constitutes a new screening instrument that is short, sensitive, and easy to use.

6.1 Introduction

The success of therapy for children with speech and language disorders depends heavily on early identification, accurate diagnosis, and the implementation of appropriate therapy (Blackman 1999). The identification of language disorders involves primary healthcare workers (e.g., health visitors, general practitioners, school nurses, and nursery staff), whereas diagnosis and therapeutic advisement involves secondary healthcare workers (i.e., multidisciplinary teams). Screening instruments create an effective link between primary and secondary healthcare practitioners. A fast, simple, sensitive, and positively predictive language-screening instrument can make a valuable contribution to the process of screening children for language problems. The detection of language problems among children between one and six years of age is particularly important, as the neurological system of speech and language develops during the first years of life (Mayeux and Kandel 1999; Mehler and Christophe 2000; Stromswold 2000). Contributing factors (e.g., mental retardation, hearing loss) —as well as the language problems themselves— should be addresses and resolved, if possible, in order to preclude further problems, including difficulties in verbal, emotional, and educational development (Coster 2001; Silva et al. 1987).

Dutch practitioners currently have no uniform instrument for screening children from one to six years of age for language problems (Luinge et al. 2002). The Health Insurance Board asked us to develop a language-screening instrument to facilitate the referral of children with possible language problems to secondary healthcare. Secondary healthcare practitioners can diagnose factors that contribute to language problems in children, including hearing disorders and psychological, psychiatric, motor, or neurological problems. It must be clear which children should be referred to secondary healthcare for the application of accurate diagnostic procedures. This article reports on the external validation of a new measurement instrument (SNEL[1]) for the early identification of speech and language disorders in the primary health care. If its test properties are satisfactory, the Groningen Public Health Service intends to implement the SNEL instrument within the primary healthcare system (Groningen Public Health Service 2004).

It is not always clear what is meant by a language problem (Stott et al. 2002; Luinge et al. 2002). Law et al. (1998) reviewed many studies about language problems and found a wide range of estimates (from 0.6% to 33.2%) for the prevalence of language problems among pre-school children. These rates are consistent with estimates provided by Dutch primary healthcare professionals, which range from one to forty percent (Luinge et al. 2002). Variation in the estimations of prevalence may be due to ambiguities in definition, cut-off scores, and the nature of language problems.

Despite ongoing discussions concerning language problems, there is consensus among researchers concerning milestones in language development (Foster 1990; Kohnstamm 1993; Krug and Mikus 1999; Kuhl 2000). A previous study showed that milestones in the language development of Dutch children from twelve to seventy-two months of age could be scaled according to the assumptions of a Mokken Item Response Model (Luinge et al. 2005c).

[1] Spraak- en taalNormen EersteLijns gezondheidszorg (speech and language norms for the primary healthcare).

The results showed that a single, unidimensional scale of diverse milestones that taps lexical, syntactic, and phonological skills, as well as both receptive and expressive language skills, is well suited for mapping progress in language development. The internal validation of the scaled language milestones (i.e., SNEL) was satisfactory. SNEL is based on the hierarchical ordering of milestones in the Dutch language, and it is expected that delays in the course of achieving language milestones are easy to detect. We further expect that primary healthcare practitioners can use such delays, in cooperation with parental report, to identify children from one to six years of age who are at risk for language problems (Law et al. 1998).

This study examines the external validation of SNEL by means of sample independence, the relationship between the scaled language milestones and age, the relationship between SNEL and age, and the relationship between SNEL and the gold standard. The study also investigates the ability of SNEL to identify children with language problems from within both the normal population and a clinical population.

For the purposes of this study, sample independence means that various subgroups of children (e.g., boys and girls, and children from different geographic regions) achieve language milestones in the same order. The advantage of a sample-independent scale is that the same ordering of milestones can be used for different subgroups, allowing levels of difficulty to be compared across subgroups (e.g., different age groups) and across repeated measurements for the same person (Mokken 1997). Repeated measurements are necessary to provide insight into the course of development, as the development of a child is a dynamic process.

The gold standard (i.e., the reference test) examines whether the scale of language milestones used in the SNEL instrument measures the construct of language and whether it is able to identify children who are at risk for language problems from within weighted samples drawn from both the normal population and a clinical population. The reference test (the Schlichting Test for Language Production) is a diagnostic test for language production, and its test properties are good (Schlichting et al. 1995). SNEL and the reference test are independent, as they assess language ability in two different ways. The reference test consists of elicitation procedures, most of which are based on imitation.

Sensitivity and specificity are basic concepts in the accuracy of screening instruments. These properties describe how well the test discriminates between children who do and do not have language problems, based on the reference test. The criteria for identifying language problems should not be too strict; they should also not exclude children who could benefit from additional assistance with language learning (Bishop 1997). To examine the relation between sensitivity and specificity, we will plot a Receiver Operating Characteristic (ROC) (Macmillan and Creelman 1991; Obuchowski 2003).

This study consists of two parts. First, it reports on the scaling results of the language milestones addressed in the previous study. Second, it examines the external validation of the SNEL language-screening instrument. The aim of SNEL is to identify language problems in children between the ages of one and six years, in order to facilitate referral to secondary healthcare.

6.2 Materials and methods

6.2.1 Scaling of milestones in language development

Participants

Data were collected from several sources. Children from one to six years of age were selected from day nurseries, playgroups, and primary schools in different parts of the Netherlands for the scaling of the milestones. Participants were further selected according to sex and specific age (e.g., a boy of one year and two months), in order to achieve a well-balanced distribution in the research population and to avoid bias (e.g., no selection based on language development). The final representative sample of 527 children consisted of 260 boys (mean = 42 months; SD = 16; range: 15–70) and 267 girls (mean = 41 months; SD = 16; range: 12–70) between the ages of twelve and seventy-two months. The children originated from a full ability range with no exclusions other than that their parents understood the Dutch language for answering the questions.

Measures

The language milestones used in the scale were derived from literature, descriptive stages, and from existing screening instruments that are based on milestones (Breeuwsma 1994; Coplan et al. 1982; Gilles and Schaerlaekens 2000; Hall 1997; Kelly and Sally 1999; Krug and Mikus 1999; Kuhl 2000; Luinge et al. 2002; Mattson et al. 2001; Mayeux and Kandel 1999; Needlman 2000; Rescorla and Alley 2001; Wachtel et al. 1994; Zuckerman et al. 1999). The milestones that were collected originate from various aspects of language comprehension and production (e.g., single-word and multi-word utterances, the naming of objects and abstract concepts, grammatical development [e.g., inflection, production of interrogative sentences], intelligibility, and the narrative capacities of young children).

A questionnaire for assessing language milestones was developed in order to determine whether children have achieved particular milestones. The questions in the questionnaire were classified into five different age groups. The researchers questioned parents or caretakers about the language development of their children by telephone. Parents were required to state only whether their children had achieved particular milestones. Dichotomous response options were chosen to facilitate interpretation. The questions were asked in Dutch.

Diagnostic statistics

A scale of several language milestones (the items) was constructed according to a nonparametric Mokken Item Response Model for dichotomous responses (Mokken 1997). The items were scaled according to the diagnostic criteria of MSPWIN 5.0, a program for Mokken scale analysis (Molenaar and Sijtsma 1982). A bottom-up strategy of automatic item-selection procedures was used to construct a scale of language milestones. The item-selection procedure started with a pair of items with the highest significant positive H coefficient and continued until the scale satisfied the scale definitions according to various diagnostic and statistical criteria.

H coefficients indicate how well items form a unidimensional construct, and they indicate the scalability of items along the scale. The following interpretation is given to the value of H: strong scale (H > 0.5), moderate scale ($0.4 \leq H \leq 0.5$), weak scale (H < 0.4) (Mokken and Lewis 1982). The dimensionality of the scale was evaluated by calculating the scalability coefficient (H).

The H^t value (i.e., the H value on a transposed data order) is a global indicator of the degree to which the ordering of the items (e.g., language milestones) is the same for every subject, according to their responses. As a general guideline, Ht must be greater than or equal to 0.30, and the percentage of negative H values should be equal to or less than ten percent of all subject responses.

The internal consistency of the scale was indicated by Rho, which can be interpreted as the IRT-based equivalent of Cronbach's alpha, which indicates how well responses to items measure an underlying construct. A scale is considered reliable if Rho is greater than or equal to 0.90 (Nunnally and Bernstein 1994).

6.2.2 External validation of SNEL

Participants

The sample of 527 participants that was used for the scaling of the milestones was also used for examining sample independence, the relationship between the scaled milestones and age, and the relationship between SNEL and age.

In addition, a weighted sample (n = 84), based on about ten percent of the 527 participants with the lowest score and about ten percent of the 527 participants, was selected at random from the normal population, in order to examine the relationship between SNEL and the reference test. The participants with the lowest scores were selected because we were particularly interested in whether low scores on the SNEL would also indicate low scores on the reference test. In total, 34 girls (mean = 46 months, SD = 16, range: 14–70) and 50 boys (mean = 42 months, SD = 15, range: 11–70) were selected from the normal population in the Netherlands.

Finally, a sample was drawn from a clinical population whose participants had been diagnosed as having language problems (11 boys and 3 girls; mean = 49 months, SD = 9, range: 33–72) and children from the weighted sample were used to examine the accuracy of SNEL (ROC) and to assess the ability of SNEL to detect children with language problems.

Measures

The Schlichting Test for Language Production was used as a gold standard (reference test) in this research. The sub-tests for sentence development and word development in the Schlichting Test have a mean reliability (coefficient α) of 0.83 (range: 0.75–0.91) and 0.85 (range: 0.71–0.76) respectively for the norm groups of 1.25 years to 6.25 years, and their correlation with other tests is high (Schlichting et al. 1995). Reliability was lowest (0.76) for the norm groups of the youngest (1;3) and the oldest children (6;3). The sub-tests are standardized for children between the ages of one year and three months to six years and three months. In this paper, participants are diagnosed as having language problems if their scores on the reference test deviate by 1.3 SD. A deviation of 1.3 SD is used in the Dutch clinical setting to diagnose language problems, and was therefore used in the validation of SNEL.

Sample independence

Sample independence was examined with reference to the sample of 527 participants. Scalability was assessed, and the applicability of the same scale to different subgroups in the population (i.e., whether items had a consistent position in the developmental sequence across different subgroups) was evaluated. Sample independence was assessed for different age groups, for boys and girls, and for children from different parts of the Netherlands.

The relationship between the scaled language milestones and age

The relationship between the separate scaled milestones as a function of age was based on the sample of 527 participants. For each scaled milestone, fractions were calculated for age groups of 6 months (12–17, 18–23, 24–29, 30–35, 36–41, 42–47, 48–53, 54–59, 60–65, and 66–72 months). Further, a figure was plotted for the ages at which each milestone was acquired by 10%, 50%, and 90% of the total number of children.

The relationship between SNEL and age

The relationship between SNEL and age was examined by plotting the SNEL scores of the participants as a function of age, based on the sample of 527 participants. A SNEL score was determined by counting the number of positive responses to the questions regarding the scaled language milestones.

The relationship between SNEL and the reference test

The weighted sample of 84 participants was used for the examination of the relationship between SNEL and the reference test. This relationship was examined by plotting the SNEL scores as a function of the raw scores on the reference test.

The relationship between sensitivity and specificity (ROC) and the accuracy of SNEL (AUROC)

Firstly, the SNEL scores of 527 participants were ranked for five age groups (12–23, 24–35, 36–47, 48–59, and 60–72 months) in order to determine percentile rankings for the SNEL scale (i.e., the 1st, 3rd, 5th, 10th, 15th, 20th, 25th, 50th, 75th, 95th, and 100th percentiles).

Secondly, the sensitivity and specificity of all of the above-mentioned percentiles were determined with reference to the weighted sample (n = 84) and the clinical sample (n = 14) (= 98 participants in total). The clinical population was also taken into account, in order to obtain additional information about the detection of the true positives (i.e., participants with language problems). Sensitivity was calculated with:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad , \tag{6.1}$$

where the symbol TP (test positives) was used to denote participants who were diagnosed as having language problems according to the reference test and SNEL, and the symbol FN (false negatives) was used to denote participants who were diagnosed as having language problems according to the reference test but not according to SNEL. A SNEL score indicates a language problem correctly when the SNEL score of a participant with a language problem is less than or equal to the examined percentile (e.g., 1st, 3rd, 5th). Specificity was calculated with:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad , \tag{6.2}$$

where the symbol TN (test negatives) was used to denote participants who were diagnosed as having language problems according to SNEL but not according to the reference test, and the symbol FP (false positives) was used to denote participants whose scores on both SNEL and the reference test indicated no language problems. A SNEL score indicates no language problem correctly when the SNEL score of a participant without a language problem is greater than the examined percentile.

Thirdly, the Receiver Operating Characteristic (ROC) was determined for all percentiles. This curve describes the sensitivity of a test (in this case, SNEL) as a function of its rate of false positives for different cut-points (the percentiles), based on the reference test (Schlichting test). The ROC curve was plotted based on the true-positive rate (i.e., the sensitivity of all examined percentiles) as a function of the false-negative rate (i.e., 1–specificity of all examined percentiles).

Finally, the area under the ROC curve (AUROC) was calculated to measure the accuracy of SNEL. The AUROC was calculated with the following formula (Macmillan and Creelman 1991):

$$A_g = 0.5 \sum (F_{i+1} - F_i)(H_{i+1} + H_i) \quad , \tag{6.3}$$

where A_g is the AUROC, H_i is the $H_{i\text{th}}$ point on the ROC curve (true-positive rate), and F_i is the $F_{i\text{th}}$ point on the ROC curve (false-positive rate). The closer the area is to 1.0, the better the test. As a general guideline, AUROC values are interpreted as follows: 0.50–0.60 = fail, 0.60–0.70 = poor, 0.70–0.80 = fair, 0.80–0.90 = good, and 90–1.00 = excellent (Tape 2004).

The ability of SNEL to detect children with language problems

The percentile with the greatest sensitivity and most satisfactory specificity was plotted based on the SNEL score as a function of the age corresponding to this percentile. Test positives (children who are correctly identified as having language problems) are more important than are test negatives (children who are falsely identified as having language problems). To show which of the participants who had been diagnosed as having language problems were situated at, below, or above the selected percentile, several cut-off scores (SD > 1 and SD > 1.3) on the reference test of participants in the weighted sample and of the participants from the clinical population were determined. The standards of SNEL were calculated using the inverse of the function for the chosen percentile.

6.3 Results

6.3.1 Scaling of milestones in language development

Table 6.1 contains eighteen scaled milestones of the SNEL scale. The scalability of the SNEL scale for children between the ages of twelve and seventy-two months was 0.87, and its reliability was 0.96. The H^t value for the entire group was 0.85; the lowest H(g) value was 0.69, and the highest H(g) value was 0.99. All diagnostic statistics were satisfactory (see Luinge et al. 2005b).

Table 6.1: Scaled milestones (18) of the SNEL-scale based on 527 children from 12 through 72 months of age (H = scalability coefficient, H^t = H on the transposed data order, Rho = reliability).

Milestones for 12 to 72 Months (n = 527)	% yes	H
1. Comprehension of 2-word sentences	98%	0.97
2. Pointing to body parts	97%	0.99
3. Production of 10 words	95%	0.99
4. Comprehension of tasks involving 3-word sentences	92%	0.99
5. Production of 2-word sentences	92%	0.99
6. Production of 3-word sentences	85%	0.96
7. Production of 3- to 4-word sentences	82%	0.93
8. About 50% intelligible	79%	0.91
9. Production of sentences using correct word order	76%	0.69
10. Naming colors	74%	0.81
11. Spontaneous storytelling	72%	0.84
12. Storytelling in response to pictures	70%	0.85
13. Production of adjectives	60%	0.88
14. About 75% intelligible	56%	0.91
15. Production of compound sentences	54%	0.85
16. Overgeneralization of verb conjugations	33%	0.77
17. About 100% intelligible	31%	0.87
18. Adult-like language production	20%	0.89

H = 0.87, Rho = 0.96, H^t = 0.85, % Negative H^t(a)-values = 0.4

6.3.2 External validation of SNEL

Sample independence

Sample independence was examined with reference to the sample of 527 participants. The applicability of the same scale (see Table 6.1) to different subgroups in the populations was evaluated. Sample independence was assessed for different age groups (12–23, 24–35, 36–47, 48–59, and 60–72 months), for boys and girls, and for participants from different parts of the Netherlands (North, East, South, and West).

The milestones with the worst sample independence were as follows: "naming colors," "sentences using correct word order," "production of adjectives," and "overgeneralization of verb conjugations." Because these items had no consistent positions in the developmental sequence across age groups, the four regions, or both, they were eliminated. A final sample-independent scale of fourteen language milestones remained (see Table 6.2). The hierarchical ordering of the fourteen remaining items was consistent for boys and girls between the ages of twelve to seventy-two months from various regions in the Netherlands.

Table 6.2: Final sample-independent scale for children from 12 through 72 months of age (H = scalability coefficient, H^t = H on the transposed data order, Rho = reliability).

Milestones for 12 to 72 Months (n = 527)	% yes	H
1. Comprehension of 2-word sentences	98%	0.97
2. Pointing at body parts	97%	0.99
3. Production of 10 words	95%	0.98
4. Comprehension of tasks involving 3-word sentences	92%	0.99
5. Production of 2-word sentences	92%	0.99
6. Production of 3-word sentences	85%	0.99
7. Production of 3- to 4-word sentences	82%	0.97
8. About 50% intelligible	79%	0.97
9. Spontaneous storytelling	72%	0.91
10. Storytelling in response to pictures	70%	0.92
11. About 75% intelligible	56%	0.91
12. Production of compound sentences	54%	0.89
13. About 100% intelligible	31%	0.95
14. Adult-like language production	20%	0.98

H = 0.95, Rho = 0.96, H^t = 0.94, % Negative H^t(a)-values = 0.5

Table 6.2 presents the final sample-independent scale (SNEL) of fourteen items, as well as the percentage of positive responses and the H coefficient for each item. The lowest H(g) value was 0.89, and the highest H(g) value was 0.99. Mokken Scale Analysis showed that the total scalability (scale H) and the reliability (Rho) of the SNEL scale were satisfactory. Scale H (0.95) was higher than 0.50, which indicates a strong scale. The H^t value of the scale (0.94) was greater than 0.30, and the percentage of negative H values was 0.5%. The time required to administer the SNEL scale varied between two and three minutes.

The relationship between the scaled language milestones and age

Figure 6.1 shows the fraction of the total number of children that had reached a scaled SNEL milestone as a function of age, based on a sample of 527 children from the normal population in the Netherlands.

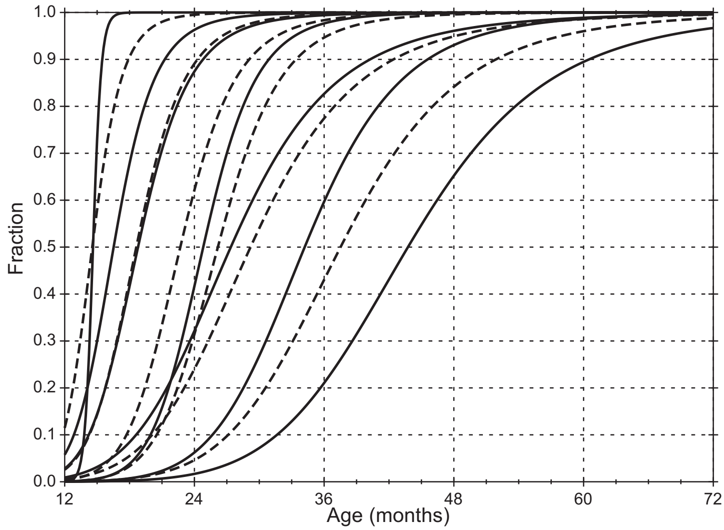

Figure 6.1: The fraction of the total number of children that acquired a scaled milestone of SNEL as a function of age, based on a sample of 527 children from the normal population in the Netherlands. For each milestone, the fraction was obtained by a fit with $1/(1+(x_0/x)^a)$ to the data points. From left to right at fraction level 0.9: 1. "Comprehension of 2-word sentences," 2. "Pointing at body parts," 3. "Production of 10 words," 4. "Comprehension of tasks involving 3-word sentences," 5. "Production of 2-word sentences," 6. "Production of 3-word sentences," 7. "Production of 3- to 4-word sentences," 8. "About 50% intelligible," 9. "Spontaneous storytelling," 10. "Storytelling in response to pictures," 11. "About 75% intelligible," 12. "Production of compound sentences," 13. "About 100% intelligible."

For each milestone, fractions were calculated for six-month age groups (12–17, 18–23, 24–29, 30–35, 36–41, 42–47, 48–53, 54–59, 60–65, and 66–72 months). In the interest of clarity, only the fitted curves are shown. Milestone 14 ("adult-like speech") could not be fitted by this function and was therefore omitted. It satisfied the assumptions of the Mokken model, however, and was therefore included in the SNEL-scale. For each milestone, these fractions were subsequently fitted with:

$$\frac{1}{1+(x_0/x)^a} \quad , \tag{6.4}$$

where x denotes age, and x_0 and a are parameters (Table 6.3). The symbol x_0 represents the age at which half of the children have achieved the corresponding milestone.

Table 6.3: Parameter values for the curves shown in Fig. 6.1.

Milestone	a	x_0 (months)
1	38.6	14.5
2	10.5	14.5
3	8.6	16.5
4	8.3	18.5
5	7.8	18.8
6	8.7	22.7
7	10.1	24.7
8	9.0	26.2
9	5.7	27.4
10	5.8	29.7
11	7.5	34.8
12	6.7	37.4
13	6.7	43.6

Figure 6.2 shows the ages at which each milestone was achieved by ten, fifty, and ninety percent of the children, based on the fitted curves shown in Fig. 6.1. The ages were obtained by intersection of the fitted curves with horizontal lines at levels 0.1, 0.5, and 0.9, respectively. These ages can be calculated with the following formula:

$$x = x_0 \left(\frac{f}{1-f} \right)^{\frac{1}{a}} \quad , \tag{6.5}$$

where x denotes age, f denotes the fraction (0.1; 0.5; 0.9), and a and x_0 the parameters as calculated above. This figure shows that the width of the distribution for a given milestone as a function of age is greater for more difficult than for less difficult milestones. The ages for the ninety-percent points are most interesting for language screening, as children who have not achieved the scaled language milestones at the ages corresponding with the ninety-percent points may be at risk for language problems. A SNEL score is based on the scores for all of the milestones included in the SNEL scale.

The relationship between SNEL and age

Figure 6.3 shows the increase in SNEL scores as a function of age, based on a sample of 527 participants. A SNEL score was determined by counting the number of positive responses to the questions regarding the scaled language milestones, as shown in Table 6.2.

The relationship between the SNEL scores and age is clearly non-linear, probably due to ceiling effects, as there is a maximum score for each age group (12–23, 24–35, 36–47, 48–59, and 60–72 months).

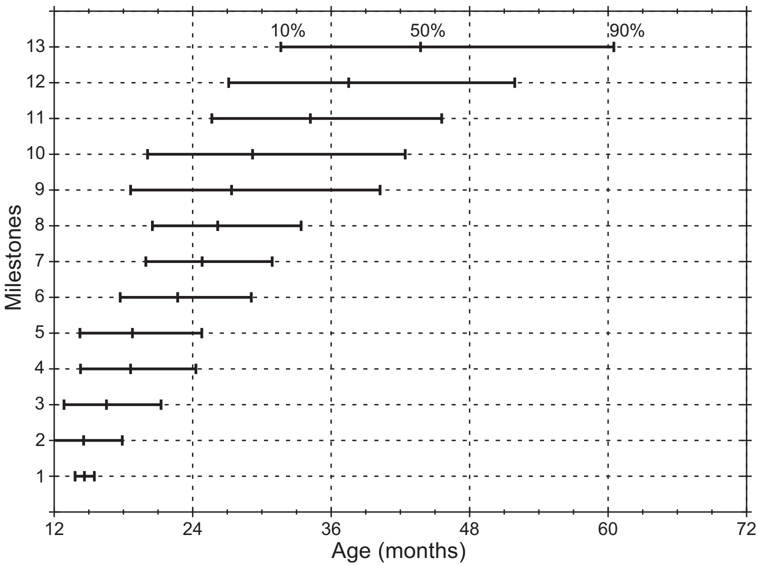

Figure 6.2: The ages at which each milestone was acquired by 10%, 50%, and 90% of the total number of children, based on the fitted curves of Fig. 6.1 vertical scale. : 1. “Comprehension of 2-word sentences,” 2. “Pointing at body parts,” 3. “Production of 10 words,” 4. “Comprehension of tasks involving 3-word sentences,” 5. “Production of 2-word sentences,” 6. “Production of 3-word sentences,” 7. “Production of 3- to 4-word sentences,” 8. “About 50% intelligible,” 9. “Spontaneous storytelling,” 10. “Storytelling in response to pictures,” 11. “About 75% intelligible,” 12. “Production of compound sentences,” 13. “About 100% intelligible.”

The relationship between SNEL and the reference test

Figure 6.4 shows the relationship between the SNEL scores as a function of raw scores on the Schlichting test, based on the weighted sample of 84 participants. The SNEL scores increased along with increases in the Schlichting scores. The relationship between the SNEL scores and age is clearly non-linear, as a consequence of the ceiling-effect caused by the maximum score on the SNEL scale, as mentioned above.

The relationship between sensitivity and specificity (ROC) and the accuracy of SNEL (AUROC)

The percentiles of the SNEL scores were determined after ranking them for each age group (12–23, 24–35, 36–47, 48–59, 60–72). Table 6.4 shows the SNEL scores for the examined percentiles (1st, 3rd, 5th, 10th, 15th, 20th, 25th, 50th, 75th, 95th, and 100th) for five age groups. The maximum scores (100th percentile) differ across age groups.

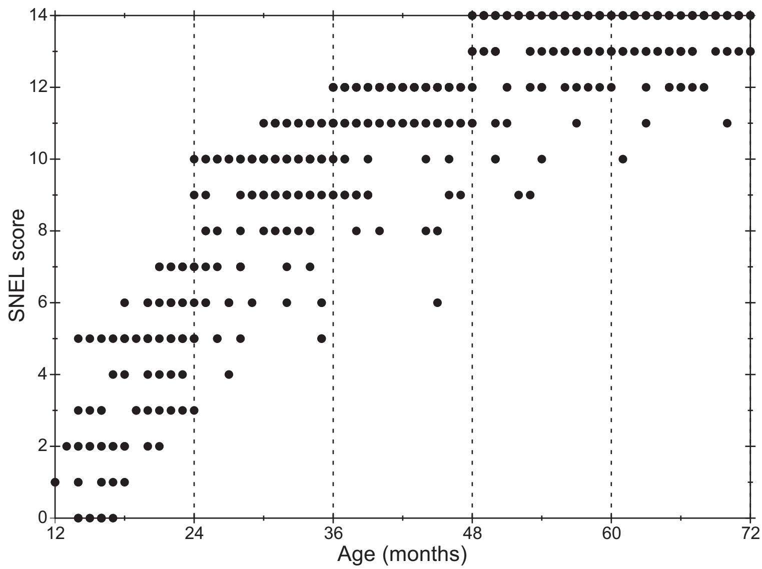

Figure 6.3: The relation between the SNEL-scores and age (months) based on a sample of 527 participants of 12 to 72 months. Only 215 points are visible, due to overlap. A SNEL score was determined by adding the yes-responses at the questions regarding the scaled language milestones of Table 6.2.

Table 6.4: The SNEL-scores for the examined percentiles for five age groups based on a sample of 527 participants.

Percentiles	12–24	24-36	36–48	48–60	60–72
1st	0	3	6	9	10
3rd	0	5	8	9	11
5th	0	5	8	10	12
10th	1	6	9	11	12
15th	1	6	10	12	13
20th	2	7	11	12	13
25th	2	8	11	13	13
50th	4	10	12	14	14
75th	5	10	12	14	14
95th	7	11	12	14	14
100th	7	12	12	14	14

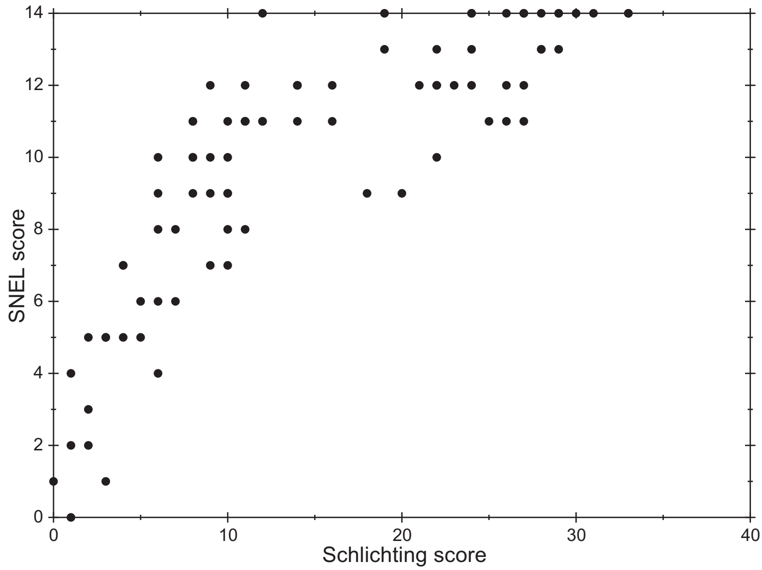

Figure 6.4: The relation between the SNEL-scores and the raw scores of the Schlichting test based on a weighted sample of 84 participants of 12 to 72 months. Only 67 points are visible, due to overlap. A SNEL score was determined by adding the yes-responses at the questions regarding the scaled language milestones of Table 6.2.

Table 6.5: The sensitivities and specificities of the SNEL-scale for the examined percentiles of Table 6.4 based on a weighted sample (n = 84) and a clinical sample (n = 14) for children in the age from 12 to 72 months. The total sample (n = 98) included 18 children with a language problem. Sensitivity was calculated by (test positives)/(test positives + false negatives) and specificity was calculated by (test negatives)/(test negatives + false positives).

Percentiles	Sensitivity	Specificity
1st	0.44	0.98
3rd	0.83	0.95
5th	0.83	0.87
10th	0.94	0.83
15th	1	0.76
20th	1	0.61
25th	1	0.5
50th	1	0.18
75th	1	0.18
95th	1	0.11
100th	1	0.04

Table 6.5 shows the sensitivity (TP /[TP + FN]; TP = test positives, FN = false negatives) and specificity (TN /[TN + FP]; TN = test negatives, FP = false positives) of the percentiles (1st, 3rd, 5th, 10th, 15th, 20th, 25th, 50th, 75th, 95th, and 100th), based on the weighted sample (n = 84) and the clinical sample (n = 14). The total sample (n = 98) included eighteen children with language problems.

Figure 6.5 shows the true-positive rate (sensitivity) as a function of the false-positive rate (1–specificity) (see also Table 6.5), as well as the fitted ROC curve for these data points. Using least squares procedures, the ROC data were fitted with:

$$y(x) = \frac{x(1+a)}{x+a} \quad , \tag{6.6}$$

where x = false-positive rate, y = true-positive rate, and a = 0.018.

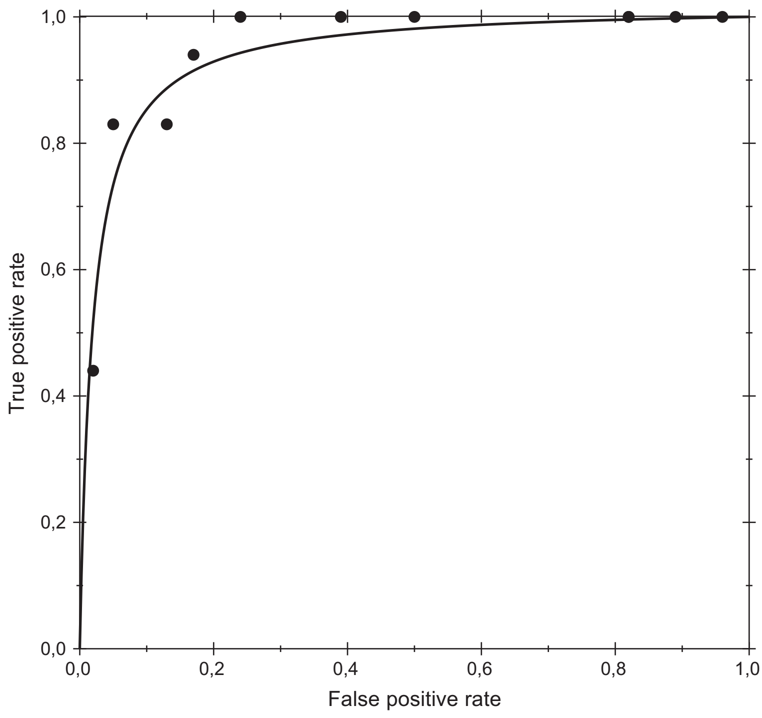

Figure 6.5: The empirical data of the true positive rate (sensitivity) as a function of the false positive rate (1–specificity) as well as the fitted ROC curve for these data points (see also Table 6.5) of the language milestones of the SNEL-scale (see Table 6.2). The data were fitted with $y(x) = (1+a)x/(x+a)$, wherein x = false positive rate and a = 0.018. Each data point of the empirical data represents a different percentile.

The assessed area under the fitted curve is 0.94, indicating that the accuracy of SNEL is very high. The area under the ROC curve can also be calculated directly from the data points as:

$$\text{AUROC} = 0.5\sum(F_{i+1} - F_i)(H_{i+1} + H_i) \quad , \tag{6.7}$$

in which F_i and H_i are the abscissa and the ordinate of the i^{th} point, respectively. This procedure yields a slightly larger value (AUROC = 0.95).

The ability of SNEL to detect children with language problems

The 10th percentile appeared to be the most appropriate percentile. It indicated high sensitivity (0.94) and sufficient specificity (0.83). In addition, this percentile is frequently used in healthcare to discriminate between children who are in the normal range of development and those who are at risk for developmental problems (e.g., with respect to weight or length).

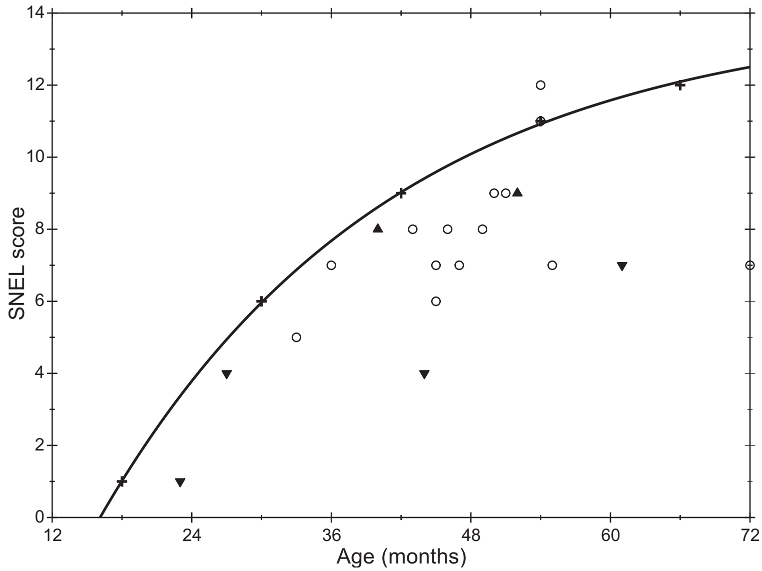

Figure 6.6: The data points (+) for the 10th percentile (see also Table 6.4) based on the sample of 527 participants in the age from 12 to 72 months. These points are fitted with $s = 14(1 - e^{-a(x-x_0)})$, wherein s = SNEL-score, a = 0.04, and x_0 = 16.1 (where x denotes age in months). Also the participants diagnosed as having a language problem (n = 6) of the weighted sample are plotted for two cut-off scores on the reference test (SD > 1: upward triangles, and SD > 1.3: downward triangles). Furthermore, the clinical population (n = 14) diagnosed as having a language problem (cut-off score SD > 1.3: circles) is inserted in the figure.

Figure 6.6 shows the data points for the 10th percentile (see also Table 6.4), the curve that was fitted to these points, and the participants from both the weighted and the clinical samples who were diagnosed as having language problems. The data for the 10th percentile were fitted with:

$$s = 14(1 - e^{-a(x-x_0)}) \quad , \tag{6.8}$$

where s denotes the SNEL score, a is 0.04, and x_0 is 16.1 (where x denotes age in months). The scores of participants from the weighted sample whose cut-off scores were SD > 1 and SD > 1.3 on the reference test (n = 6) were inserted into the calculations illustrated in Fig. 6.6, in order to determine whether SNEL should also detect participants with low-average scores (SD > 1). Only one participant from the clinical population (SD > 1.3) had a SNEL score above the 10th percentile (false negative). The SNEL scores of all of the other participants who had language problems were at or beneath the 10th-percentile curve, which indicates that these participants would have been detected by SNEL.

Table 6.6 shows the SNEL scores that should be achieved at specific ages, according to the curve fitted to the 10th-percentile data presented in Fig. 6.6. The rounded-off ages are determined using the inverse of the function for the fitted line presented in Fig. 6.6:

$$x = \frac{1}{a} \ln\left(\frac{14}{14-s}\right) + x_0 \quad , \tag{6.9}$$

where x is age (months), s is SNEL score, a denotes 0.04, and x_0 denotes 16.1. The symbol "ln" represents the natural logarithm. For example, Table 6.6 shows that a SNEL score less than or equal to 2 for a twenty-two-month-old child indicates a possible language problem. This SNEL score is situated at or beneath the 10th-percentile curve presented in Fig. 6.6. Figure 6.7 shows the SNEL score as a function of age for the 10th percentile (data points from Table 6.6) ($s = 14(1 - e^{-a(x-x_0)})$).

Table 6.6: SNEL-scores that should be achieved at several ages according to the curve in Fig. 6.6 fitted to the 10th percentile, based on a sample of 527 participants.

Age (months)	SNEL-score
12–17	0
18–19	1
20–21	2
22–24	3
25–26	4
27–29	5
30–32	6
33–36	7
37–41	8
42–46	9
47–54	10
55–64	11
65–81	12
82	13

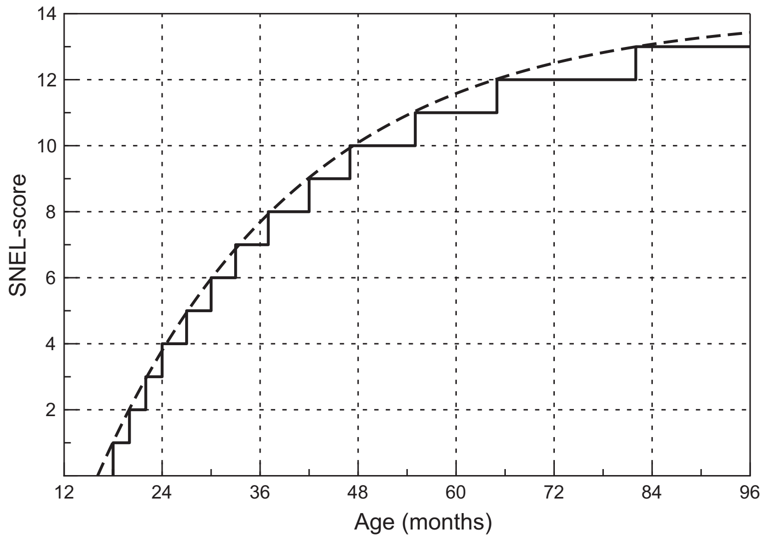

Figure 6.7: The SNEL score as a function of age for the 10th percentile (data points from Table 6.6) ($s = 14(1 - e^{-a(x-x_0)})$).

6.4 Discussion

This study has shown SNEL to be an accurate language-screening instrument that is sensitive enough to detect language problems in children from twelve to seventy-two months of age. The unidimensional and sample-independent SNEL scale constitutes a new instrument that is short and easy to use. Scores can be derived from parental report. The hierarchical ordering in the achievement of language milestones is consistent for boys and girls between the ages of twelve and seventy-two months from various regions in the Netherlands. The scaled milestones can therefore make a valuable contribution to the development of a universal screening instrument by translating them into other languages, testing them for different populations, and validating them according to different reference tests and clinical judgments.

Furthermore, the use of parental report is very useful for the screening of young children, as the language of these children usually refers to concepts that are found within the home environment (e.g., "daddy book"). Artificial test situations may therefore tend to underestimate children's language abilities. With parental report, it is not necessary to involve children in the screening. This feature facilitates the completion of the screening, as it removes the necessity of scheduling and transporting children. Finally, parental report may allow the assessment of intra-individual variability in language development (Van Dijk et al. 2001), as the reports are based on longer periods of time.

The internal validation showed that both the scalability and the reliability of SNEL were good. In addition, the external validation of SNEL yielded satisfactory results. As expected, SNEL scores increased with age. Furthermore, the standards for SNEL were investigated according to percentiles and not according to the mean SNEL scores of each age group. The mean scores may not have reflected language development accurately, as the ceiling effects that were mentioned earlier could have lowered the standards for the SNEL scale. Percentiles are less sensitive to ceiling effects. Furthermore, high SNEL scores (and the accompanying ceiling effects) are not interesting for screening purposes, as language-screening instruments should detect only children with low SNEL scores, which indicate possible language problems.

The correspondence between the outcomes of SNEL and the reference test permits the conclusion that SNEL and the reference test measure the same construct, language production. The reference test was validated for children of fifteen months of age and older (Schlichting et al. 1995), whereas the present study includes children from the age of twelve months. For this reason, the external validation of SNEL for the youngest children should be interpreted with caution. In addition, the score of one participant from the clinical population (n = 14) who was diagnosed as having a language problem was ambiguous. This score fell on the 10th percentile, as shown in Table 6.5 (age = 54 months, SNEL score = 11), but it was still within the range of the data points.

The use of the ROC curve in this study also provided an elegant method that avoided some of the problems that are associated with other methods (Mitchell 2003). Another advantage of the use of ROC curves is that they allow straightforward conclusions to be drawn from the positions of the data points (Obuchowski 2003).

This study provides neither estimates of the prevalence of language problems nor predictive values, as the weighted sample (n = 84) from the normal population includes five percent of the lowest SNEL scores and a sample at random. Despite this weighted sample, however, relatively few of the children (5.7%) had possible language problems. The prevalence of language problems among the participants in this study falls within the range of the prevalence estimates developed by Law et al. (1998) (range 0.6–33.2%) and by the Dutch primary healthcare professionals (range 1–40%) (Luinge et al. 2002).

In conclusion, the unidimensional and sample-independent SNEL scale constitutes a new measurement instrument that is short and easy to administer. It appears sufficiently sensitive to detect children who are at risk of language problems. The results of both the internal and external validations were good. The scalability of SNEL was strong (H = 0.95), its reliability was high (Rho = 0.96), and the concurrent accuracy measured against parental report of language development for children between the ages of twelve and seventy-two at months was excellent (AUROC = 0.94).

Limitations of this study and further research

SNEL is intended as a screening instrument for use in primary healthcare. It consists of yes-or-no questions, and it can be completed by parental report. Administration of the scale takes about three minutes. Although the psychometric qualities of SNEL are promising, SNEL requires further examination in future research. The reference test is supposed to measure language production. Other language tests (e.g., language comprehension tests) should be able to indicate whether SNEL also screens for other aspects of language development. In addition, because the test sample in this study was representative of the Dutch population, little can be said about the applicability of the results to other languages or to language problems that are related to psychiatric disorders (e.g., autism). Another limitation of this study is that the external validation was based on a relative small sample. A larger sample from the normal population as well as the clinical population should provide more evidence for the external validation results of this study. Additionally, the predictive values of SNEL should be determined based on a large sample from the normal population.

The applicability of SNEL as a tool for primary healthcare practitioners is currently under examination by the Groningen Public Health Service. They are also evaluating whether SNEL is cost-effective as a language-screening instrument (Groningen Public Health Service 2004).