University of Groningen

Identifying Milestones in Language Development for Young Children Ages 1 to 6 Years

Visser-Bochane, Margot I; Reijneveld, Sijmen A; Krijnen, Wim P; van der Schans, Cees P;
Luinge, Margreet R

Published in:
Academic pediatrics

DOI:
[10.1016/j.acap.2019.07.003](https://doi.org/10.1016/j.acap.2019.07.003)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Visser-Bochane, M. I., Reijneveld, S. A., Krijnen, W. P., van der Schans, C. P., & Luinge, M. R. (2020). Identifying Milestones in Language Development for Young Children Ages 1 to 6 Years. *Academic pediatrics*, 20(3), 421-429. <https://doi.org/10.1016/j.acap.2019.07.003>

Copyright

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

PII: S1876-2859(19)30333-X
 DOI: <https://doi.org/10.1016/j.acap.2019.07.003>
 Reference: ACAP 1383

To appear in: *Academic Pediatrics*

Received date: 21 March 2019
Accepted date: 7 July 2019

Please cite this article as: Margot I. Visser-Bochane MSc , Sijmen A. Reijneveld MD, PhD , Wim P. Krijnen PhD , Cees P. van der Schans PT, CE, PhD , Margreet R. Luinge PhD , Identifying Milestones in Language Development for Young Children Ages 1-6 Years, *Academic Pediatrics* (2019), doi: <https://doi.org/10.1016/j.acap.2019.07.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Identifying Milestones in Language Development for Young Children Ages 1-6 Years

Margot I. Visser-Bochane, MSc¹; Sijmen A. Reijneveld, MD, PhD²; Wim P. Krijnen, PhD¹; Cees P. van der Schans, PT, CE, PhD^{1,2,3}; Margreet R. Luinge, PhD^{1,4}

Affiliations:

¹ Research Group Healthy Ageing, Allied Health Care and Nursing, Hanze University Groningen, Applied Sciences, Groningen, the Netherlands

² Department of Health Sciences, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

³ Department of Rehabilitation Medicine, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

⁴ Department of Otorhinolaryngology, Head & Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Corresponding Author: Margot I Visser-Bochane, Research Group Healthy Ageing, Allied Health Care and Nursing, Hanze University of Applied Sciences Groningen, Petrus Driessenstraat 3, 9714 CA Groningen; e-mail: m.i.visser-bochane@pl.hanze.nl

Keywords: language development; prevention; developmental language disorder; child health surveillance; language acquisition

Running head: Identifying milestones of language development

Word counts: Abstract: 160 words; main text: 3144 words

Acknowledgements The authors thank the participating teachers, health nurses, speech-language therapy students, and parents.

Financial disclosure: This work was supported by the Netherlands Organization for Health Research and Development (ZonMw; grant number 200330002).

Declaration of interest The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. All authors read, edited, and contributed to the manuscript, and approved submission of the final version.

Abbreviations:

IRF-Item Response Function

IRT-Item Response Theory

Abstract

Objective: Language development is important for children's success in life. Therefore, language is monitored by child health care professionals and parents, but a uniform set of milestones in language development is lacking. Our aim was to identify a set of clear and distinctive milestones that empirically reflect language development in children aged one to six years of age.

Methods: We obtained a community-based sample of 1,381 parents reporting on milestones derived from clinical signs and currently used language screening instruments. We used nonparametric Item Response Theory analysis to identify milestones belonging to one unidimensional scale.

Results: Twenty-six milestones were excellently scalable (item H coefficients 0.62-0.90) and formed a strong scale (total H coefficient 0.83). The final set of identified milestones covered vocabulary, grammar, and communication, with an item ordering that holds for all children.

Conclusion: This unidimensional set of 26 clear and distinctive milestones reflects language development in young children and can be used as instrument to monitor language development.

What's New

We empirically identified a set of 26 milestones in vocabulary, grammar, and communication belonging to one scale that reflects language development in children aged one to six. The milestones are feasible for monitoring language development in community-based settings.

Introduction

Language development is a basic skill for all children, and essential for participation in everyday life, more specifically: a child's social and emotional development, and educational success.¹⁻³ Atypical language development has a major impact on a child and its environment, and needs to be identified as early as possible.⁴ Therefore, language development of children is monitored with various instruments by professionals working in child health care and professionals working in educational settings, and by parents.⁴⁻⁷ In order to adequately identify atypical language development, clear milestones of typical language development are required. Various milestones, e.g. the well-known 'puts two words together at age two', are used for early assessment. However, empirical evidence lacks on the best milestones for identification of language development from age 1 to 6 years.

Milestones in language development should represent the underlying construct of language development, including its intertwined aspects such as speech, language, and communication. Speech involves the production of speech sounds. Children start babbling from the age of 6 months, next speech sounds develop until the age of 7 years.⁸ Language involves the comprehension and production of words, sentences, and ideas. The first words emerge when a child is about 12 months, up from that age vocabulary increases.⁹ Next, grammar development starts when children put two words together, typically occurring around a child's second birthday.¹⁰⁻¹¹ Thereafter, various elements of grammar, such as syntax and morphology develop.¹² Communication involves nonverbal and verbal use of language in interaction with other persons, and is seen as a necessary condition for language development.¹³ Up till now, it is unclear whether milestones in speech, language, and communication reflect the same underlying construct, that is language development.

Children develop language with great variability in the onset time and rate of vocabulary, and grammar, e.g. mean length of utterance, proportion of verbs in total

vocabulary, and onset of use of irregular nouns and verbs.^{12,14,15} Because of the variation in onset time and rate, it is difficult to describe average language development that holds for all children.^{14,15} However, it may be possible to identify more general milestones in speech, language, and communication that have the same order for all children. To our knowledge, two previous studies identified a general set of ordered language milestones.^{11,16} These milestones covered language in different ways, namely, in the study of Luinge et al¹⁶ as vocabulary, grammar, and speech, but not communication, and in the study of Sheldrick et al¹¹ as a broader construct than language, namely vocabulary, grammar, and communication but not speech, and motor and cognitive development in addition. Also, milestones in language development are used in developmental screening questionnaires, such as the widely implemented Ages and Stages Questionnaire (ASQ)¹⁷. In contrast with Luinge and Sheldrick, constructors of the ASQ did not identify ordered milestones for the whole continuum of language development. Instead, the ASQ includes six milestones on language development that partially vary for 21 specific age windows (ages 1 to 5.5 years). These milestones regard comprehension and production of words and sentences, also grammar, but not speech and communication (use of language in interaction). Further research is necessary to find out what general milestones reflect the construct of language development.

Furthermore, such general milestones should be feasible for parents and professionals in well-child care and education in order to monitor language development. Therefore, we need to identify clear milestones that can be observed by parents and professionals. A set of clear distinctive milestones that have the same order for all children can contribute to the empirical description of typical language development, and to the monitoring of language development by parents and by professionals working in community-based settings. Therefore, the aim of our study was to identify a set of clear distinctive milestones that empirically reflect language development in children from one to six years of age.

Methods

Study Design

We conducted a cross-sectional study to identify a set of milestones (items) that reflect language development in children from 1 to 6 years of age, and to assess the 1) feasibility, concerning use of the items for a certain age range; 2) the comprehensibility of the items for parents; and 3) the construct validity of the resulting set. The construct validity of the set was evaluated by interpreting the results of Mokken scale analysis that builds upon nonparametric Item Response Theory. The consecutive steps for the identification of the set of milestones that reflect language development are shown in Figure 1.

The study was approved by the Medical Ethical Committee of the University Medical Center of Groningen, Netherlands (number NL45253.042.13), and registered at trialregister.nl (number 5746).

Statistical approach – Mokken scaling Procedure

A way to identify a set of clear distinctive items that reflect an underlying ability such as language development is Mokken scaling,¹⁸ which belongs to nonparametric Item Response Theory (IRT). The identified items as identified in a Mokken scaling procedure belong to one underlying ability and are individually distinctive for a specific range of the ability (Figure 2). The advantage of Mokken analysis is that items can be identified that contribute to the rank ordering of persons on the same underlying ability,¹⁹ where the identified items reflect the development of language. Mokken analysis is not the only IRT model. Other frequently used models are parametric IRT-models, such as the Rasch model. Rasch analysis provides logistic item characteristic information, highly similar to Mokken. Moreover, the basic IRT assumptions are similar for Mokken and Rasch. However, the Rasch framework does not include an automated item selection procedure or a general procedure to

check invariant item ordering. It consists of an iterative estimation procedure to fit the logistic item characteristic curves. This works only fine for a selected set of items sharing the same measurement properties, making the Rasch model unsuited for the selection of items.

Sample

We used two-step sampling to recruit parents/caregivers of children from one to six years of age. First, we took a random selection of municipalities. Second, from these, we randomly selected daycare centers (to include the parents of 1-2-3 year old children) and kindergartens (to invite parents of 4-5 year old children). We also recruited well-child clinics throughout the Netherlands (1-2-3 year olds). By recruiting children via daycare centers and well-child clinics for the 1-2-3 year olds, and kindergartens for the 4-5 year olds, we obtained a community-based sample representative for the general population in the Netherlands. There were no exclusion criteria. Parents provided written informed consent.

Measures and Procedure

We constructed parent questionnaires for the ages 1-2, >2-3, >3-4, >4-5, >5-6 years to describe language development based on items derived from various sources. Candidate items were derived from clinical signs for atypical language development,²⁰ and from all currently used Dutch screening instruments: Van Wiechen²¹; Speech and Language Norms for Primary Healthcare (SNEL)¹⁶; the questionnaire on speech and language problems from the Dutch national association for the education of children with speech and language problems retrieved from <http://www.simea.nl>; and from a Dutch well-child clinic retrieved from <https://www.ggddrenthe.nl/scholen/logopedie>. During the data collection, all items were phrased in Dutch as yes/no questions on language ability and, if applicable, supplemented with an example. Face validity (item is specific, measurable, relevant, clear for parents) of the items was discussed by eight professionals in child language, resulting in an initial set of 75

items in total. The items were administered in Dutch. We used forward-only translation of the original items to English for international publication. Development of the instrument was finalized by dividing the items over five partially overlapping questionnaires, each of which pertained to one age group (for example: *'Does your child ask questions with 'why'?'* was asked in the age groups 3, 4, and 5-year olds but not in the age groups 1, and 2 year olds).

We measured feasibility as whether the items were appropriate for a certain age range according to the response of parents on the items and whether all of the items were appropriate for use in daily practice. Comprehensibility was measured by transcribing parents' reactions to the items other than a 'yes' or 'no' answer, and asking them whether they actually understood each of them.

We further measured background characteristics of the child and family. These concerned age, gender, birthweight, length of pregnancy, language situation at home, and highest level of education achieved by the mother. Frequently reported risk factors for language disorders are gender, birthweight, length of pregnancy, educational level of the mother⁵. Therefore, we have included these factors as background variables. Education level was classified into three categories: low (primary school or less, and pre-vocational education), middle (secondary education), and high (higher vocational education and university).

We collected data on language and background characteristics from the parents using a telephone interview of, on average, 20 minutes that was conducted by speech and language pathology students who were specifically selected and trained for this study.

Analyses

First, we described the background characteristics of the sample. Second, we assessed feasibility by calculating the proportion of 'yes' and 'no' answers of parents on the items for

the entire group. Items that answered 100% 'yes' were removed. Third, we assessed the comprehensibility of the items. Parental responses on the items, other than 'yes' or 'no', were transcribed and coded followed by group discussions per item by three examiners regarding specificity and clarity for parents. Items that appeared multi-interpretable were then labeled as not meeting the criteria (i.e. not specific or not clear for parents) and were therefore removed from the item set.

Fourth, we identified items and investigated the construct validity of the identified items by comparing the results of Mokken scale analysis with the assumptions of the Mokken model, viz scalability, monotonicity and item ordering. For the missing observations, we applied a substitution by the item mean rounded to the nearest integer. Non-discriminating items with proportions equal to one were omitted from further analysis. We then applied Mokken scale analysis to identify items and evaluate the (1) scalability; (2) monotonicity; and (3) item ordering based on the responses on the items by the parents. We assessed scalability by evaluating unidimensionality and homogeneity. Unidimensionality was investigated using an automated item selection procedure (AISP), taking a lower bound of 0.3 for homogeneity and significance level alpha 0.05.^{22, 23} Homogeneity was quantified with Loevinger's coefficient for the total scale (H) and each item (H_i). The scale coefficient H is a weighted mean of the quality of the items (H_i s). We adopted the generally accepted interpretation on scalability as: $0.3 \leq H < 0.4$ weak, $0.4 \leq H < 0.5$ moderate, and $0.5 \leq H < 1.0$ strong.²² (2) We assessed monotonicity by evaluating the item response function (IRF). A monotonically non-decreasing IRF indicated a positive relation of an item to the latent trait of language development. Items violating monotonicity were removed. We evaluated (3) item ordering for all pairs of items with a backward selection procedure.²² From the item pair with the largest number of ordering violations, we removed the item with the lowest H coefficient. The procedure identifies a set of items without such violations. All items were plotted

(proportions of children with a ‘yes’-score for each item against total scores on the final set of milestones), to visualize the relationship between the indicator as a function of language development. The steepness of the plotted curves indicates the distinctiveness of an item, with steeper slopes indicating better discrimination. Moreover, the set of items should cover the developmental period. Last, we determined the age of acquisition for each item based on fitted curves, for percentages of the children passing of 15%, 50% and 85%. Data were analyzed using R version 3.3.1 using the Mokken-package version 2.7.7.²²

Results

Sample

We collected data on 1,381 children attending 146 daycare centers, 128 kindergartens, and 10 well-child clinics. The first 150 parents of children from one to six years of age (73 boys, 77 girls) were assigned to the sample for feasibility analysis. Next, a sample consisting of 1,231 parents of children ranging from one to six years of age was used in the analysis of the comprehensibility, and construct validity. Characteristics of our sample are presented in Table 1. The sample was representative for the Dutch population with respect to birthweight, pregnancy duration, and educational level of the mother, three factors closely related to the outcomes. For comparison: 6.2% of the Dutch children is born with a birthweight below 2500 grams, 7.2% of the children is born preterm, and 56% of the females (25-35 years old) has a higher educational level, vs 4.4% low birthweight and 5.9 % preterm (Volksgezondheidszorg.info, 2019), and 60% high educational level (Statistics Netherlands, 2019) in our sample. Also, age groups did not differ on birthweight and pregnancy duration. However, one way analysis of variance showed that age groups slightly differed regarding educational level [$F(4,1223) = 2.70, p = 0.029$]. Specifically, two year olds had the highest

percentage of higher educated mothers (64.2%), and three year olds had the lowest percentage of higher educated mothers (53.3%).

Feasibility and comprehensibility

We removed 22 out of 75 items for the following reasons: 100% yes-responses (100% no-responses did not occur), i.e. the item measures an ability that all children already have acquired (eight items), the item has no learning curve and is thereby not specific for an age group (eight items), the items have duplicate meaning (four items), and the item did not fit the construct (two items). The remaining items were distributed over five overlapping questionnaires corresponding to the five age groups. Because some items were assigned to other age groups, we began a new sample for analysis of comprehensibility and psychometric properties. The analysis of the comprehensibility resulted in the exclusion of five items (Supplemental Table A) because these items were not clear to parents.

Construct validity

The responses on the remaining 48 items were generally of good quality. On the 98 (0.17%) missing observations on items over all of the respondents, we applied a substitution to the nearest integer of the column mean. Five items were excluded because they appeared to be too easy, having an item proportion of endorsement of .995 or higher (Supplemental Table A).

Testing scalability (unidimensionality) on the remaining 43 items with automated item selection procedure (AISP) resulted in the selection of 40 items. Three items were unscalable, meaning that they did not fit the dimension of the selected 40 items, and had to be excluded (Supplemental Table 1). The set of 40 items had a total H coefficient of 0.77. The items showed strong scalability with item coefficients H_i ranging from 0.51 to 0.91. The lowest item H coefficient (0.51) regarded the item *'Is the pronunciation of your child (with*

exception of the letter r) good?', and the highest H coefficient (0.91) regarded the item *'Does your child understand tasks consisting of two words? For example: 'coat on' or 'look there'.'*

The assessment of monotonicity revealed no violation by any of the items, therefore, no items were excluded. That implied that the probability of a positive response for each item increased with a more extensive language ability of the child. Investigating invariant item ordering by backward selection resulted in 15 steps and thereby the exclusion of 14 items (Supplemental Table A). The frequencies of violations are reduced by subsequent removal of the worst item until no violations on item ordering were found. This procedure resulted in a set of items where all of the children have the same ordering of items.

Final set of items

The final set of items consisted of 26 items regarding milestones in vocabulary, grammar, and communication, reflecting language development covering the age range from one to six. The items were ordered covering language development (expressed in the total score on the final set of items), with item curves that increased (ascending slopes) at increasing levels of language development (Figure 3). The item H coefficients were between 0.62 and 0.90 which indicated excellent scalability (Supplemental Table B). The scale H coefficient was 0.83 (SE=0.01) which indicated a strong scale. Items proportion of endorsement ranged from 0.36 to 0.99, i.e., 36 to 99% of the full sample of children had acquired the language ability that was involved. A percentage of 99% indicates that the corresponding item does not discriminate well over all age groups, but rather only in the younger age groups. The age of acquisition for each item based on fitted curves showed increasing ages of acquisition for the various items. Relatively many items were easy, i.e. feasible for younger children, and fewer were difficult, i.e. only suitable for older children.

Only four items (item 23-26) covered the full language development of children aged 40 to 71 months (Figure 4).

Discussion

In this study, we identified 26 milestones in one dimension that reflect typical language development in children from one to six years of age. All milestones belonged to one, unidimensional, scale that covers vocabulary, grammar, and communication, with an item ordering that holds for all children. All identified milestones were feasible for the age range (1-6) and were comprehensible for parents.

The identified milestones reflect language development in children from one to six years of age, belonging to one scale, covering vocabulary, grammar, and communication, and that does not include the production of speech sounds. To our knowledge, only two previous studies also identified a set of milestones in language development on a unidimensional scale,^{11,16} but none of these covered vocabulary, grammar, and communication in a concise way. The removal of all milestones on the production of speech sounds is consistent with the idea that speech and language should be considered to be different dimensions.²⁴ Our study empirically demonstrates that milestones in vocabulary, grammar, and communication form a unidimensional scale that reflects language development in children aged one to six, and that production of speech sounds is not part of that unidimensional scale for children aged one to six.

We found an ordering of milestones that holds for all included children, with sentence comprehension (item 3,4,5) preceding sentence production (e.g. item 11,15). This ordering aligns with the current knowledge on language acquisition that comprehension precedes production.⁹ It has also been suggested that the inverse pattern could occur in some children or for detailed linguistic forms, e.g. object pronouns, subject-verb-object word order.²⁵ Our

findings suggest that these deviations at least do not occur if assessing a general level of comprehension or production using parental report. Only such more general milestones are more useful for community-based use as detailed linguistic forms require professional knowledge about linguistics and experimental assessment strategies. Our study thus shows that on a more general level that parents perceive that language comprehension indeed precedes language production.

All milestones in the final set were comprehensible for parents supporting the previous conclusion that language development can be monitored by parents when using easily observable items.^{4, 11, 16, 26, 27} The easiness of observation can be attributed to our process of item formulation that resulted in comprehensible, and positively formulated milestones in language development. We used no negatively formulated milestones, which may have minimized response bias as reversing some milestones has been shown to result in errors due to inattention or confusion by the respondent.²⁸ Moreover, milestones did not focus on deficits which could raise unnecessary concerns. It is thus feasible to measure language development in children from one to six years of age derived from parental observation.

In contrast with other studies,^{29, 30} we found that items on comprehension (item 3, 4, 5) could be quantified based on parent report. In our study, the included comprehension items consist of short tasks such as 'Does your child understand tasks consisting of two words? For example: 'coat on' or 'look there''. Parents can judge whether their child understands this, because they can observe the current response behavior of the child.³¹ Other studies assess the comprehension based on a checklist with words checked by the parents, e.g. the MacArthur Bates Communicative Development Inventory Words and Gestures (CDI-WG)³² and the Language Development Survey (LDS).³³ The validity of such checklists with words is questioned, because it is difficult for parents to judge whether a child really understands a word. For example, the words in the checklists may be used at home, but that does not

necessarily mean the child comprehends these words.³⁰ Yet, assessing comprehension is important as deficits in language comprehension are a clear indicator for persisting language disorders.^{9,34} Our study shows that language comprehension can be measured based on parent-report regarding the performance of short tasks by their child that show that the child understands the task.

The scale has been developed in Dutch, but is likely to be of use in other Germanic languages, such as English and German, as the items represent milestones in language that are very common in related languages and are not restricted to the Dutch language, e.g. ‘says two-word sentences’ (item 11). Use of the scale in other languages requires a language specific verification of the presented norms.

A major strength of our study is our use of a large, community-based sample. In addition, we had a very small number of missing values, and prevented information bias by using trained interviewers. Our study also had some limitations, the most important one being that high educated mothers were slightly overrepresented in our sample for two-year old children, i.e., 64% compared to the national average being approximately 56% (Statistics Netherlands, 2019). This may have slightly affected the estimated ages of acquisition as children of these parents have been shown to have slightly faster language development.³⁵

Conclusions

In conclusion, our findings show that language development in children aged one to six can be measured with one scale. Our final set of empirical milestones encompasses milestones in vocabulary, grammar, and communication and reflects increases in one underlying construct of language development in children aged one to six. The properties of this scale, i.e., consisting of parent-reported, clear and distinctive milestones, make the set feasible as well as promising for monitoring language development in community-based settings. The scale was constructed for quick community based developmental screening. The

ages of attainment can be used as indicative norms of typical language development during well child visits. However, cut-off norms for screening purposes can best be determined against a gold standard. Our study has several implications for further research. First, the final scale should be validated in order to establish its test characteristics in terms of sensitivity and specificity, and predictive values in routine care. Second, the use of the scale in community settings deserves further study regarding feasibility and added value in routine care. Last, further research is needed to determine whether this unidimensional scale can also be used in other languages.

References

1. Boyle CA, Decoufle P, Yeargin-Allsopp M. Prevalence and health impact of developmental disabilities in US children. *Pediatrics*. 1994;93(3):399-403.
2. Conti-Ramsden G, Durkin K, Mok PL, Toseeb U, Botting N. Health, employment and relationships: Correlates of personal wellbeing in young adults with and without a history of childhood language impairment. *Soc Sci Med*. 2016;160:20-28.
3. Roulstone S, Law J, Rush R, Clegg J, Peters T. Investigating the role of language in children's early educational outcomes. Research Report on Department for Education, UK; 2011
4. Bright Futures Steering Committee, & Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*. 2006;118(1):405-420.
5. Berkman ND, Wallace I, Watson L, Coyne-Beasley T, Cullen K, Wood C, Lohr KN. Screening for speech and language delays and disorders in children age 5 years or younger. A Systematic Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 120. AHRQ Publication No. 13-05197-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2015.
6. Williams C. Teacher judgements of the language skills of children in the early years of schooling. *Child Lang Teach Ther*. 2006;22(2):135-154.
7. Glascoe FP, Dworkin PH. The role of parents in the detection of developmental and behavioral problems. *Pediatrics*. 1995;95(6):829-836.
8. Dodd B. *Differential diagnosis and treatment of children with speech disorder*. John Wiley & Sons; 2013.

9. Simonsen HG, Kristoffersen KE, Bleses D, Wehberg S, Jørgensen RN. The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*. 2014;34(1):3-23.
10. Conti-Ramsden G, Durkin K. Language development and assessment in the preschool period. *Neuropsychol Rev*. 2012;22(4):384-401.
11. Sheldrick RC, Perrin EC. Evidence-based milestones for surveillance of cognitive, language, and motor development. *Acad Pediatr*. 2013;13(6):577-586.
12. Jiang H, Logan JA, Jia R. Modeling the Nature of Grammar and Vocabulary Trajectories From Prekindergarten to Third Grade. *J Speech Lang Hear Res*. 2018;61(4):910-923.
13. Kuhl PK. Early language acquisition: cracking the speech code. *Nat Rev Neurosci*. 2004;5(11):831.
14. Bates E, Dale PS, Thal D. Individual differences and their implications for theories of language development. *The handbook of child language*. 1995;96-151.
15. Fenson L, Dale PS, Reznick JS, et al. Variability in early communicative development. *Monogr Soc Res Child Dev*. 1994;i-185.
16. Luinge MR, Post W, Goorhuis-Brouwer SM. The language screening instrument SNEL. *Educational and Child Psychology*. 2007;24:20–30.
17. Squire J, Twombly E, Bricker D, Potter L. ASQ-3: User's guide. Baltimore: Paul H. Brookes Publishing Co.; 2009.
18. Sijtsma K, Molenaar IW. *Introduction to nonparametric item response theory* (Vol. 5). Sage, 2002.
19. Meijer RR, Baneke JJ. Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychol Methods*. 2004;9(3):354.

20. Visser- Bochane MI, Gerrits E, van der Schans CP, Reijneveld SA, Luinge MR. Atypical speech and language development: a consensus study on clinical signs in the Netherlands. *Int J Lang Commun Disord.* 2017;52(1):10-20.
21. Laurent de Angulo MS, Brouwers-de Jong EA., Bijlsma-Schlösser JFM, Bulk-Bunschoten AMW, Pauwels JH, Steinbuch-Linstra I. *Ontwikkelingsonderzoek in de jeugdgezondheidszorg; het Van Wiechenonderzoek—De Baecke-Fassaert Motoriekttest.* Assen: Van Gorcum; 2005.
22. Van der Ark LA. Mokken scale analysis in R. *J Stat Softw.* 2007;20(11):1-19. URL <http://www.jstatsoft.org/v20/i11/>.
23. Van der Ark LA. New developments in Mokken scale analysis in R. *J Stat Softw.* 2012;48(5):1-27.
24. Shriberg LD, Tomblin JB, McSweeny JL. Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *J Speech Lang Hear Res.* 1999;42(6):1461-1481.
25. Hendriks P, Koster C. Production/comprehension asymmetries in language acquisition. *Lingua.* 2010;120(8):1887-1897.
26. Miller LE, Perkins KA, Dai YG, Fein DA. (2017). Comparison of parent report and direct assessment of child skills in toddlers. *Res Autism Spectr Disord.* 2017;41: 57-65.
27. Sachse S, Von Suchodoletz W. Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatr.* 2008;29(1):34-41.
28. Van Sonderen E, Sanderman R, Coyne JC. Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one.* 2013;8(7):e68967.
29. Feldman HM, Dollaghan CA, Campbell TF, Kurs- Lasky M, Janosky JE, Paradise JL. Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Dev.* 2000;71(2):310-322.

30. Tomasello M, Mervis CB. The Instrument is great, but measuring comprehension is still a problem. *Monogr Soc Res Child Dev.* 1994;59(5):174-179.
31. Dale PS, Bates E, Reznick JS, Morisset C. The validity of a parent report instrument of child language at twenty months. *J Child Lang.* 1989;16(2):239-249.
32. Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick JS. *MacArthur-Bates Communicative Development Inventories: user's guide and technical manual*. PB Brookes; 2014
33. Rescorla L. The Language Development Survey: A screening tool for delayed language in toddlers. *J Speech Hear Disord.* 1989;54:587–599.
34. Vehkavuori SM, Stolt S. Screening language skills at 2;0. *Infant Behav Dev.* 2018;50:174-179
35. Rice ML, Hoffman L. Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2; 6 to 21 years of age. *J Speech Lang Hear Res.* 2015;58(2):345-359.

Table 1. Characteristics of the sample used for the identification of the milestones.

	12 - 23 months (n=215)	24 - 35 months (n=303)	36 - 47 months (n=263)	48 - 59 months (n=239)	60 - 72 months (n=211)	Total (n=1,231)
Characteristics	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	
Gender						
Male	110 (51.2)	160 (52.8)	143 (54.4)	115 (48.1)	107 (50.7)	635 (51.6)
Female	105 (48.8)	143 (47.2)	120 (45.6)	124 (51.9)	104 (49.3)	596 (48.4)
Birthweight (grams)*						
High (>5000)	0 (0.0)	1 (0.3)	1 (0.4)	2 (0.8)	1 (0.5)	5 (0.4)
Average (2500-5000)	206 (98.1)	276 (93.6)	242 (94.2)	225 (95.4)	196 (95.6)	1145 (95.2)
Low (<2500)	4 (1.9)	18 (6.1)	14 (5.4)	9 (3.8)	8 (3.9)	53 (4.4)
Pregnancy duration						
Serotine (>42 weeks)	11 (5.1)	17 (5.6)	13 (4.9)	17 (7.1)	12 (5.7)	70 (5.7)
Term (37-42 weeks)	198 (92.1)	260 (85.8)	229 (87.1)	208 (87.0)	193 (91.5)	1088 (88.4)
Preterm (<37 weeks)	6 (2.8)	26 (8.6)	21 (8.0)	14 (5.9)	6 (2.8)	73 (5.9)
Language						
Monolingual	183 (85.1)	249 (82.2)	224 (85.2)	210 (87.9)	180 (85.3)	1046 (85.0)
Bi-, Multilingual	32 (14.9)	54 (17.8)	39 (14.8)	29 (12.1)	31 (14.7)	185 (15.0)
Education of mother*						
High	135 (62.8)	194 (64.2)	139 (53.3)	145 (60.6)	129 (61.1)	1228 (60.4)
Middle	73 (34.0)	92 (30.5)	99 (37.9)	79 (33.1)	73 (34.6)	416 (33.9)
Low	7 (3.2)	16 (5.3)	23 (8.8)	15 (6.3)	9 (4.3)	70 (5.7)

* Numbers do not always add up to $N = 1,231$ due to missing data; Birthweight ($n = 28$), Education of mother ($n = 3$).

Figure 1. Consecutive steps in the process of identification of milestones.

Figure 2. Mokken scale model for language development. An ordering of distinctive items (milestones) that hold for all children reflecting the same underlying language development.

Figure 3. The fitted proportion of positive responses of the parents on each item (y-axis) by the total score on the final set of 26 items, i.e. increasing language development (x-axis). Each curve represents a single item.

Figure 4. Ages at which each of the 26 milestones is achieved for 15% (light grey), 50% (grey), and 85% (dark grey) of the children.

Figure 1

Figure 2

Figure 3

Figure 4