DTU Library

A Danish Sentence Corpus for Assessing Speech Recognition in Noise in School-Age Children

Koiek, Shno; Nielsen, Jens Bo; Kjærbæk, Laila; Gormsen, Maria Baltzer; Neher, Tobias

Published in:
Trends in Hearing

Link to article, DOI:
10.1177/2331216520942392

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Trends in Hearing
Volume 24: 1–8

DOI: 10.1177/2331216520942392
journals.sagepub.com/home/tia
SAGE

A Danish Sentence Corpus for Assessing Speech Recognition in Noise in School-Age Children

Shno Koiek[1,2], Jens Bo Nielsen[3], Laila Kjærbæk[4], Maria Baltzer Gormsen[5], and Tobias Neher[1,2]

Abstract

For the audiological assessment of the speech-in-noise abilities of children with normal or impaired hearing, appropriate test materials are required. However, in Denmark, no standardized materials exist. The purpose of this study was to develop a Danish sentence corpus suitable for testing school-age children. Based on the 600 validated test sentences from the Danish DAT (*D*agmar, *A*sta, or *T*ine) corpus, 11 test lists comprising 20 sentences each were carefully constructed. These lists were evaluated in terms of their perceptual similarity and reliability with a group of 20 typically developing, normal-hearing children aged 6 to 12 years. Using stationary speech-shaped noise and diotic stimulus presentation, speech recognition thresholds (SRTs) were measured twice per list and participant at two separate visits. The analyses showed that six test lists were perceptually equivalent. These lists are characterized by a grand average SRT of −2.6 dB signal-to-noise ratio, a test–retest improvement of 0.6 dB, and a within-subject standard deviation of 1.1 dB signal-to-noise ratio. The other lists were characterized by slightly higher SRTs, slightly larger training effects, and slightly larger measurement uncertainty, but were otherwise also usable. Overall, it is therefore concluded that the developed corpus is suited for assessing speech recognition in noise in Danish 6- to 12-year olds. The corpus is publicly available.

Keywords

speech test, speech recognition in noise, pediatrics

Received 24 February 2020; revised 16 June 2020; accepted 22 June 2020

Children are often exposed to noise (e.g., in classrooms), which causes difficulties with speech recognition (e.g., Shield & Dockrell, 2003). In general, the adverse effects of noise on speech recognition diminish as children get older, but problems can be observed across childhood (Johnson, 2000; Soli & Sullivan, 1997; Werner & Boike, 2001). Consequently, reliable methods for assessing speech recognition in noise in children are essential, especially when deficits are suspected. In Germany, the 'Oldenburger Kinder Satztest' was developed for that purpose (Wagener et al., 2006). The Oldenburger Kinder Satztest consists of three-word pseudo-sentences that include a numeral, an adjective, and a noun (e.g. "four red flowers"). It has been found to be usable with children aged four and above (Neumann et al., 2012; Wagener et al., 2006). Another speech test that was designed for the use with children is 'FreeHear,' which is available in British English (Moore et al., 2019). In this test, three spoken digits are presented against a background of babble noise. Because of its simple sentence structure, this test has been found to be suitable for children as young as 4 years.

[1]Institute of Clinical Research, University of Southern Denmark
[2]Research Unit for Oto-Rhino-Laryngology, Odense University Hospital
[3]Hearing Systems, Technical University of Denmark
[4]Department of Language and Communication, University of Southern Denmark
[5]Department of Audiology, Odense University Hospital

Corresponding Author:
Tobias Neher, Institute of Clinical Research, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark.
Email: tneher@health.sdu.dk

In Denmark, a number of speech materials are available for clinical and research purposes, for example, DANTALE I (Elberling et al., 1989) and DANTALE II (Wagener et al., 2003). DANTALE I includes lists of monosyllabic words for the measurement of discrimination scores. There are 4 lists containing 20 words each, which are considered suitable for children aged 5 and above. Furthermore, there is one list containing 20 words for younger children. However, none of these lists has been formally validated and standardized. DANTALE II includes lists of semantically unpredictable five-word sentences consisting of a name, a verb, a numeral, an adjective, and a noun each (e.g. "Kirsten bought four red flowers"). These lists were evaluated with normal-hearing adults, revealing significant training effects.

Another available Danish speech test is the conversational language understanding evaluation (CLUE) test (Nielsen & Dau, 2009). The sentence material used for the CLUE test stems from a large database, which includes everyday conversational sentences from Danish newspapers, magazines, books, and so on. The principles and procedure behind the CLUE test stem from the Hearing in Noise Test (HINT; Nilsson et al., 1994). In 2010, Nielsen and Dau (2011) developed a Danish HINT that was based on the speech material from the CLUE test. The Danish HINT consists of 10 test lists and 3 practice lists containing 20 sentences each. It has been shown to be a reliable measure of speech recognition for Danish adults with normal or impaired hearing (Nielsen & Dau, 2011). Although the reliability of the CLUE and Danish HINT has not been investigated with children, it is likely to be lower due to the properties of the speech material. First, the HINT material contains many words that are not part of the vocabulary of young children. Second, the sentences in the Danish HINT have different grammatical structures, some of which may be too complicated for children to understand. Third, the length of the sentences (five words) may exceed the memory capacity of younger children.

Another available set of Danish sentences is the DAT corpus, which is an open-set, low-context, multitalker speech corpus (Nielsen et al., 2014). The DAT corpus includes 600 unique sentences that have a fixed, simple structure. More specifically, they all make use of a carrier sentence that starts with a female name (i.e., *D*agmar, *A*sta, or *T*ine; hence, the abbreviation DAT) and that contains two short keywords. During testing, listeners are only required to repeat the two keywords per sentence. Because the keywords are located immediately before the last word of the carrier sentence ("i går"), measurements made with the DAT material are unlikely to be affected by memory issues. In addition, the low predictability of the sentences makes them difficult to memorize. As part of the validation of this material, the sentences were systematically distributed into 30 test lists of approximately equal intelligibility. The lists were then validated using speech recognition measurements performed with normal-hearing and hearing-impaired Danish adults. The developed test material was found to produce consistent speech recognition thresholds (SRTs) across the different test lists and to be sensitive to interindividual differences (Nielsen et al., 2014).

To summarize, there are currently no standardized Danish materials that are suited to assessing speech recognition in noise in children. The purpose of this study was to address this shortcoming. In particular, the aim was to develop a set of test lists suited for 6- to 12-year olds. The aforementioned properties of the DAT corpus motivated us to use this material as the basis for the development of a 'child-friendly' DAT corpus called the børneDAT corpus. Ideally, this corpus should be characterized by small training effects, high test list equivalence, and low measurement uncertainty. In this study, we assessed these aspects by performing test–retest measurements with the created test lists over a time period of approximately 1 to 2 weeks.

Materials and Methods

Ethical approval for this study was obtained from the Regional Committees on Health Research Ethics for Southern Denmark.

Generation of Test Lists

For the compilation of the test lists, the 600 validated test sentences from the DAT corpus (Nielsen et al., 2014) were used. As pointed out above, all of these sentences have a fixed, simple structure. That is, they start with Dagmar (D), Asta (A), or Tine (T) and contain two short, neutral, and concrete keywords (nouns), for example, "Dagmar tænkte på *en teske* og *en næse* i går" ("Dagmar thought about *a teaspoon* and *a nose* yesterday"). In terms of semantic properties, the noun pairs are not related, which makes them impossible to predict. For each name, there are 200 test sentences that were uttered by 1 of 3 professional female talkers. The average fundamental frequencies of these talkers are 176 Hz (Dagmar), 188 Hz (Asta), and 243 Hz (Tine), respectively. During recording, all the sentences were articulated with a neutral voice without any accentuation. Furthermore, the talkers were instructed to maintain a constant level throughout and a speech rate similar to that of everyday conversational Danish. Mispronounced words, unintended noises, unnaturally long pauses between some words and silent periods at the beginning and the end of the sentences were removed.

For this study, 220 sentences with keywords belonging to the vocabulary of a typical 6-year-old were selected. As part of the selection process, 2 audiological researchers and 1 researcher in the language development of children (i.e., three of the authors) individually went through all 600 sentences of the DAT corpus and assessed them in terms of their suitability for testing 6- to 12-year olds. Those sentences which all 3 researchers judged to be suitable were kept and combined into 11 lists containing 20 sentences each. In the original DAT study, the intelligibility of each sentence was assessed in a listening experiment involving 16 normal-hearing adult participants (Nielsen et al., 2014). The assessment was based on the assumption that sentence intelligibility can be quantified in terms of the signal-to-noise ratio (SNR) at which both keywords are correctly identified, and that more intelligible sentences are recognized at lower SNRs than less intelligible sentences. In the current study, the original sentence intelligibility assessments were assumed to be also valid for school-age children. Using the procedure from the original DAT study, sentences with relatively high and low intelligibility were counterbalanced at the beginning of each test list, while sentences with approximately equal intelligibility were used toward the end of each list (see Nielsen et al., 2014 for details). For a given list, only sentences uttered by the same talker (and thus starting with the same name) were used. In this manner, four D-lists (D1, D2, D3, and D4), three A-lists (A1, A2, and A3), and four T-lists (T1, T2, T3, and T4) were created. The 11 compiled test lists are provided in the Appendix.

Participants

A total of 20 typically developing, normal-hearing children (13 females) participated in the study. They were aged 6 to 12 years (mean: 8.7 years). Their parents provided written informed consent, and the children received a gift card at the end of the study.

All participants fulfilled the following inclusion criteria: (a) normal middle-ear function, (b) pure-tone hearing thresholds $\leq$ 25 dB HL at all standard audiometric frequencies from 125 to 8000 Hz, (c) normal speech discrimination in quiet, (d) native Danish speakers, (e) normal language development, and (f) normal cognitive function. Otoscopy and tympanometry were performed to examine the outer and middle ears of all participants. Children with any type of obstruction or infection in the ear canal and/or type-B or type-C tympanograms were excluded. Standard pure-tone audiometry was carried out using supra-aural headphones. Next, speech discrimination in quiet at the most comfortable level was tested using the DANTALE I material (Elberling et al., 1989). Listeners with discrimination scores <90%-correct were also excluded. Language development of the children was assessed using the Peabody Picture Vocabulary Test (Dunn et al., 1965), which confirmed normal language skills for all participants. Cognitive development was assessed based on parental reports. In addition, a custom-made questionnaire was administered that included questions related to the child's mother tongue, whether the child was monolingual, and the income and education level of the child's parents. All participants in this study were monolingual, native Danish speakers and came from families with middle to high incomes.

Apparatus and Procedures

All measurements were conducted in a soundproof booth. To evaluate the 11 created test lists in terms of their perceptual similarity and reliability, SRT measurements were made. The speech stimuli were presented diotically in stationary speech-shaped noise through supra-aural, free-field equalized headphones (Sennheiser HDA200). The speech-shaped noise was talker-specific. The order of the test lists was balanced across the participants. The starting level of the speech signal was 67 dB SPL. The level of the noise was fixed at 60 dB SPL. The equipment was calibrated using a 01 dB FUSION sound level meter and a GRAS 43AA-S2 CCP ear simulator kit. The SRTs were measured using the adaptive procedure from the Danish HINT (Nielsen & Dau, 2011). The participants were verbally instructed to repeat the two keywords in each sentence. In case of any doubts, they were encouraged to guess. Responses were scored as correct if both keywords were repeated accurately.

Before the start of the actual measurements, all participants performed one SRT measurement in quiet and two SRT measurements in noise. The lists used for these purposes were training lists from the original DAT material. A short break was included after the first five SRT measurements and whenever a participant felt tired. A set of retest measurements was made on average 10 days (range: 5–19 days) after the first set of measurements.

Statistical Analyses

The collected data were analyzed using SPSS (IBM) version 25. To begin with, the consistency of each child's test–retest data was assessed based on squared difference scores and scatter plots. This resulted in the data of one child being excluded from all subsequent analyses because of clear inconsistencies. To verify the equality of variance in the datasets, Levene's test was used. This showed equality of variance in the datasets for all test lists. Shapiro–Wilk's test, normal *Q*–*Q* plots, and box plots were applied to examine the distribution of the collected data. The results showed that all but two SRT datasets (i.e., those corresponding to the A1- and A2-lists collected at the second visit) were normally distributed. The two nonnormal distributions were due to a single outlier in each case. Exclusion of the concerned data points resulted in normal distributions for all datasets. To assess the influence of test list, visit and talker, repeated-measures analyses of variance (ANOVAs) were performed. In all cases, a significance level of 5% was used.

Results

Perceptual Validation of the Generated Test Lists

Figure 1 shows the mean list SRTs for the two visits. The grand average SRT across all lists and participants was −2.2 dB SNR with an across-subject standard deviation (*SD*) of 1.5 dB SNR. The within-subject *SD* was 1.2 dB SNR. The list-mean-SRT *SD* was 0.6 dB SNR, and the maximum deviation from the overall mean was 1.1 dB.

Given that the sentences of the D-, A-, and T-lists were uttered by three talkers with different voice characteristics, the SRT measurements were analyzed in terms of a potential talker effect. The overall mean SRTs of the D-, A-, and T-lists were found to be −1.5 dB SNR, −2.8 dB SNR, and −2.3 dB SNR, respectively. A one-way ANOVA comparing the mean SRTs of the D-, A-, and T-lists showed a significant effect, $F(2, 195) = 21.4$, $p < .001$. Post hoc comparisons based on Tukey's test revealed that the mean SRT of the D-lists was significantly higher than those of the A- and T-lists, whereas the mean SRTs of the A- and T-lists did not differ from each other (see Table 1).

To investigate the perceptual similarity of the A- and T-lists, a two-way repeated-measures ANOVA with the within-subject factors visit and list was carried out. This showed statistically significant effects of list, $F(6, 102) = 5.3$, $p < .001$, and visit, $F(1,17) = 18.0$, $p = .001$. Post hoc comparisons based on Tukey's test showed that the T1-list differed significantly from the T2-, T3-, and A-lists (all $p < .05$). In contrast, the A1-, A2-, A3-, T2-, T3-, and T4-lists did not differ from each other (all $p > .05$).

To investigate the perceptual similarity of the four D-lists and the T1-list, another two-way repeated-measures ANOVA was carried out. This showed statistically significant effects of list, $F(4, 68) = 4.2$, $p = .004$, and visit, $F(1, 17) = 14.7$, $p = .001$. Post hoc comparisons based on Tukey's test showed that the D2-list differed significantly from all the other lists (all $p < .05$). In contrast, the D1-, D3-, D4-, and T1-lists did not differ from each other (all $p > .05$).

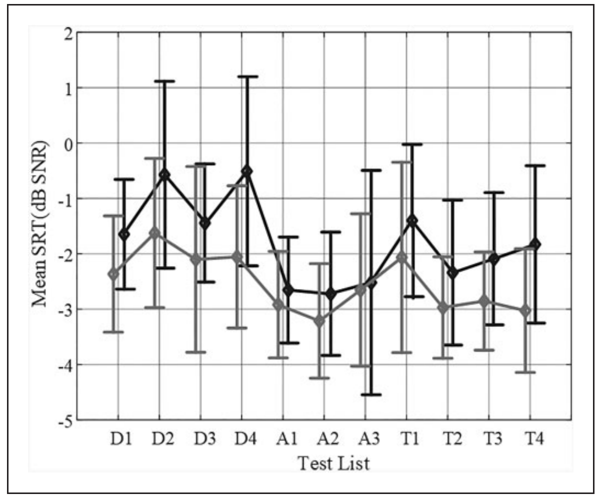

Figure 1. Mean List SRTs for the First (Black) and Second (Gray) Visit. Error bars show ±1 *SD*. SRT = speech recognition threshold; SNR = signal-to-noise ratio.

Table 1. Results of Post Hoc Tests Comparing the Mean SRTs of the D-, A-, and T-Lists.

Talker 1	Talker 2	Mean SRT difference (dB)	*p*
T	A	0.5	.06
D	T	0.8	<.001
	A	1.2	<.001

Note. SRT = speech recognition threshold.

Definition of Training and Test Lists

Based on the statistical results above, two sets of lists were defined: one for training (D1, D3, D4, and T1) and one for testing (A1, A2, A3, T2, T3, and T4). Figure 2 shows the mean SRTs of the lists for the two sets. For the six test lists, the grand average SRT was −2.6 dB SNR, the average test–retest improvement 0.6 dB, the within-subject *SD* 1.1 dB SNR, and the *SD* of the list means 0.2 dB SNR. For the four training lists, the grand average SRT was –1.7 dB SNR, the average test–retest improvement 0.9 dB SNR, the within-subject *SD* 1.3 dB SNR, and the *SD* of the list means 0.3 dB SNR.

Age Effect

Given that the study participants covered a relatively wide age span (6–12 years), the effect of age on the SRT results was also tested. Figure 3 shows a scatter plot of age against the grand average SRT (calculated across all 11 test lists). As expected, older children achieved lower (better) SRTs compared with younger children. The relationship between age and mean SRT was statistically significant, $r(18) = −.50$, $p < .05$.

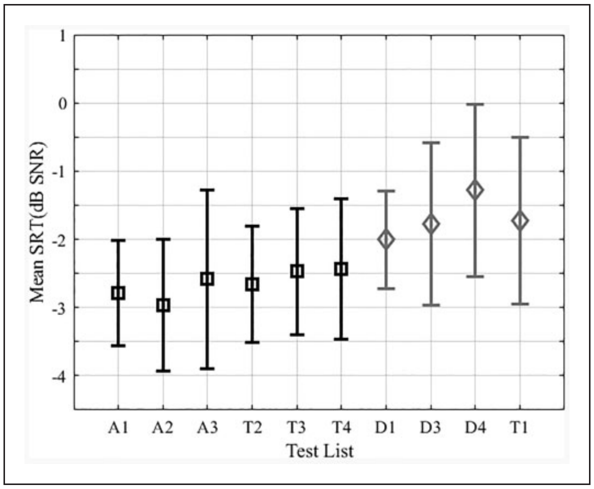

Figure 2. Mean SRTs for the Two Sets of Perceptually Equivalent Test Lists (Test Lists: Black Squares; Training Lists: Gray Diamonds). Error bars show ±1 *SD*. SRT = speech recognition threshold; SNR = signal-to-noise ratio.

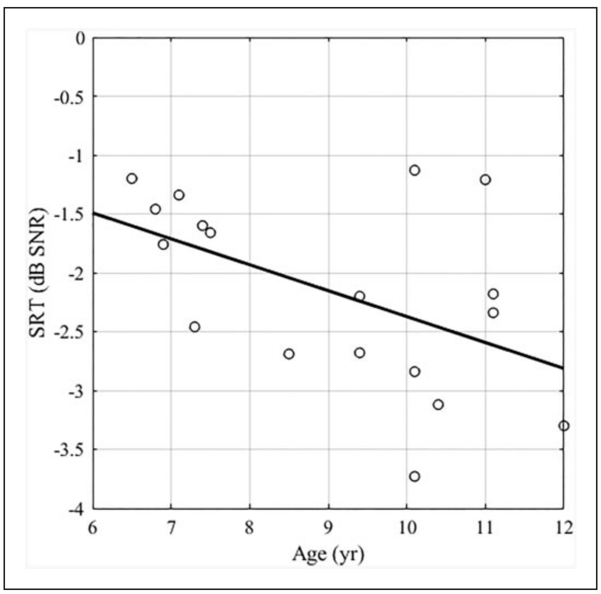

Figure 3. Scatter Plot of Age Versus Grand Average SRT With a Least-Squares Regression Line. SRT = speech recognition threshold; SNR = signal-to-noise ratio.

Discussion

The current study aimed to develop a Danish speech material, which is suitable for assessing speech recognition in noise in school-age children. More specifically, the objective was to develop a set of test lists with small within-subject and between-list variation that can be used for performing SRT measurements with 6- to 12-year olds. Eleven lists comprising 20 sentences each were compiled based on the validated sentence material from the DAT corpus (Nielsen et al., 2014). Test list equivalence and measurement reliability were examined with the help of 20 typically developing, normal-hearing Danish children.

To assess the properties of the børneDAT material, the obtained results were compared with those for the Danish HINT (Nielsen & Dau, 2011) and DAT material (Nielsen et al., 2014). In this study, the grand average SRT of the 11 developed lists was −2.2 dB SNR, which is comparable to the mean SRT of the Danish HINT obtained with 16 normal-hearing adults (–2.5 dB SNR; Nielsen & Dau, 2011). The within-subject *SD* for these lists was 1.2 dB SNR, which is somewhat larger than that of the Danish HINT (0.9 dB SNR). The larger within-subject *SD* could be due to the large difference in age of the participants of the two studies (children vs. adults) as well as differences in the speech materials used (e.g., without vs. with context). Another possibility could be the larger variability in terms of language skills for the children tested here compared with the adults tested for the Danish HINT.

Since the sentences of the D-, A-, and T-lists were uttered by three different talkers, we considered the influence of talker on the results. We found that the D-lists resulted in significantly higher mean SRTs than the A- and T-lists. As part of the development of the original DAT corpus, the effect of talker was also considered. For SRTs measured with adult listeners and a speech-on-speech masking paradigm, Nielsen et al. also observed higher SRTs for talker D. Overall, these findings suggest that the D-sentences are slightly less intelligible than the A- and T-sentences. Nielsen et al. (2014) did not investigate the cause of this, but a likely explanation is differences in pronunciation between the talkers.

Ideally, the test lists of a given speech corpus should result in very similar SRT measurements, so they can be used interchangeably. Based on our results, the developed material includes six lists (A1, A2, A3, T2, T3, and T4) with equivalent mean SRTs that we propose to use for test purposes. The mean test–retest improvement for these lists was 0.6 dB, corresponding to that observed for the Danish HINT. The within-subject *SD* of 1.1 dB SNR across these lists was only slightly larger than the 0.9 dB SNR observed for the Danish HINT with adult listeners. Thus, the reliability of the two materials is similar. Furthermore, four other lists (D1, D3, D4, and T1) had also equivalent mean SRTs. The mean SRT for these lists (−1.7 dB SNR), mean test–retest improvement (0.9 dB), and within-subject *SD* (1.3 dB SNR) were slightly higher/larger than for the other set, which is why we recommend to use the second set for training purposes. In terms of their list-mean-SRT *SD*s, the two sets of lists were very comparable (0.2 and 0.3 dB SNR, respectively) and also very similar to that of the Danish HINT (0.3 dB SNR). The maximum deviation from the overall mean SRT was only 0.4 dB, which is smaller than that observed for the Danish HINT (0.6 dB). In general, this indicates high equivalence of the developed test lists with respect to the measured SRT.

Regarding the effects of age, this study confirmed that younger children struggle more to understand speech in noise compared with older ones. This is consistent with a large body of research comparing the effects of noise on younger versus older children (e.g., Johnson, 2000; Soli & Sullivan, 1997; Walker et al., 2019; Werner & Boike, 2001), and can be traced back to developmental changes in terms of language skills of school-age children (e.g., Firmansyah, 2018). At a more general level, this finding implies that since younger children understand speech especially poorly in noisy environments it is particularly important to provide classrooms with good acoustical conditions for them.

The developed speech material is publicly available and can be used in speech-based research with Danish 6- to 12-year olds. Nevertheless, some limitations should also be noted. First, to allow the developed speech material to be used in clinical practice, age-specific normative data need to be collected. Second, the validation we performed was restricted to normal-hearing children. Future work will therefore also examine the properties of the speech material with children with hearing loss.

Conclusions

Based on the validated sentence material from the Danish DAT corpus, a new set of sentences aimed at Danish 6- to 12-year olds was constructed and evaluated in terms of its reliability. Six test lists were found to be equivalent in terms of their mean SRTs. A small training effect was observed, suggesting that the lists can be reused after 1 to 2 weeks. Four lists considered to be suitable for training purposes resulted in mean SRTs that were on average 1 dB higher than the mean SRTs of the other six lists, but otherwise equivalent and usable. Overall, the developed material seems therefore suitable for research studies with Danish 6- to 12-year olds.

Appendix

Developed Test and Training Lists

Test Lists. **Test list A1:** “Asta tænkte på . . . og . . . i går”

1. en kaptajn, en saks; 2. en pingvin, en streg; 3. en træsko, et skjold; 3. en fisk, en knæskal; 4.en skovtrold, en mark; 5. en biograf, en fælde; 6. et arbejde, en paryk; 7. en plade, en byfest; 8. en weekend, en gaffel; 9. en hytte, et papskilt; 10. en zebra, en børste; 11. en boble, et græskar; 11. en ost, et postkort; 12. en vandpyt, en tønde; 13. et vindue, en jagthund; 14. en bogreol, en konge; 15. en port, en paraply; 16. en vask, en flodhest; 17. en pelikan, en suppe; 18. en isbjørn, en kvinde.

Test list A2: “Asta tænkte på . . . og . . . i går”

1. en ketsjer, en tånegl; 2. en reklame, en kok; 3. en diamant, en cykel; 4. et lotteri, en lygte; 5. en skovsø, en dåse; 6. en klub, et bolsje; 7. en svamp, en kagedej; 8. en solsort, en potte; 9. en bjørn, en fodbold; 10. et eventyr, en duks; 11. et batteri, en klods; 12. en opgave, et spark; 13. et hjerte, en prins; 14. en hårtot, en støvle; 15. et paprør, en æske; 16. et smykke, en elefant; 17. en trøje, en bageovn; 18. et tebrev, et telt; 19. en kjole, et blåbær; 20. en due, et posthus.

Test list A3: “Asta tænkte på . . . og . . . i går”

1. en brand, et værksted; 2. en færge, en gulvklud; 3. et kamera, en stork; 4. en busk, en undulat; 5. en tyr, en kantsten; 6. en garage, en tærte; 7. en tavle, en vaffel; 8. en salat, en skramme; 9. en cirkel, et hylster; 10. et blad, et korthus; 11. en guldfisk, en kost; 12. en terning, en skøjte; 13. en fletning, en cigar; 14. en form, et dørskilt; 15. en småkage, en rejse; 16. et stempel, en smil; 17. et dyr, en kridtstreg; 18. en vinter, en storby; 19. en pige, en trompet; 20. en vikar, en sandal.

Test list T2: “Tine tænkte på . . . og . . . i går”

1. et træhus, en sovs; 2. et æble, en krabbe; 3. et styr, en grøntsag; 4. en landsby, en balje; 5. en tepose, en hest; 6. et postbud, en kant; 7. en gang, et fjernsyn; 8. en tur, en limstift; 9. et æsel, en tromme; 10. et krybdyr, en pande; 11. en dessert, en fløjte; 12. en flagstang, en hvalp; 13. et skakspil, en abe; 14. en skammel, en dagbog; 15. en ble, en bjergtop; 16. en ølkasse, en giraf; 17. et kødben, en stemme; 18. et fortov, en frakke; 19. en drøm, en sløjfe; 20. en regnskov, en heks.

Test list T3: “Tine tænkte på . . . og . . . i går”

1. en finger, en skovtur; 2. et hus et, vinglas; 3. et spring, en fortand; 4. et kor, et fodspor; 5. et træ, en krystal; 6. en kanon, en stolpe; 7. et tog, et spædbarn; 8. en sportsmand, et bord; 9. en ballon, et stopur; 10. en præmie, en klasse; 11. en skorsten, et føl; 12. en påfugl, en buket; 13. en sværd, en tulipan; 14. en kran, en delfin; 15. et palads, en dyrlæge; 16. et pindsvin, et flag; 17. en agurk, et vandløb; 18. en bagdør, en fridag; 19. et håndtryk, en ørken; 20. en kommode, en frugt.

Test list T4: “Tine tænkte på . . . og . . . i går”

1. en sportsvogn, en sild; 2. en uniform, et bælte; 3. et apotek, et sving; 4. et vejskilt, en tiger; 5. en fe, en jordklump; 6. et drivhus, en fugl; 7. et juletræ, en tand; 8. en tekop, et fnis; 9. en hamster, en prøve; 10. en grøft, en vulkan; 11. en hveps, en fyrste; 12. en kålorm, en ferie; 13. en lasso, et bakspejl; 14. et apparat, en dyne; 15. en snog, et græsfrø; 16. en dans, en lastbil; 17. en parasol, en stige; 18. en ske, et stålrør; 19. en tråd, et skumbad; 20. en skov, en tegning.

Training Lists. **Training list D1:** "Dagmar tænkte på ... og ... i går"

1. en teske, en næse; 2. en madkurv, en kogle; 3. et legehus, en rift; 4. en gorilla, en datter; 5. en sytråd, en strand; 6. en pels, en kontakt; 7. en klovn, en høstak; 8. en trappe, en tilbud; 9. en smørklat, en vest; 10. en bilkø, en vifte; 11. en fad, en blinklys; 12. en gård, et græsstrå; 13. en troldmand, en kasse; 14. en torsk, en bande; 15. et soltag, en sauna; 16. en bro, et fængsel; 17. et net, en kronprins; 18. en plads, et kapløb; 19. en pære, et kort-spil; 20. en film, en knivspids.

Training list D3: "Dagmar tænkte på ... og ... i går"

1. en sprøjte, et job; 2. en brandmand, en flue; 3. en nabo, en maskine; 4. en vejkant, en mose; 5. en kiks, en jungle; 6. et knaphul, en krog; 7. en skorpe, en rygsæk; 8. en malkeko, et svar; 9. en skat, et kæledyr; 10. en glass-kål, en sut; 11. en sok, en sporhund; 12. en sommer, en fedtplet; 13. en snabel, et skilt; 14. en elastik, en gnist; 15. en tabel, et spil; 16. en fest, et bilhorn; 17. en pejs, et snobrød; 18. en grotte, et betræk; 19. en økse, en run-dkreds; 20. en ispind, en stribe.

Training list D4: "Dagmar tænkte på ... og ... i går"

1. et rensdyr, et slot; 2. en sæk, et sejlskib; 3. en frisør, en klokke; 4. et bål, en kænguru; 5. et hindbær, en mus; 6. en festsal, en rist; 7. en krølle, en strømpe; 8. en kamel, en hilsen; 9. en gråspurv, en seng; 10. en båd, et spring-vand; 11. en mursten, en kappe; 12. en gårdsplads, en mønt; 13. en lakrids, en doktor; 14. en ankel, en rovfugl; 15. en tryne, en blyant; 16. en ridetur, en billet; 17. en sko, en popsang; 18. en adresse, et krus; 19. en formand, en synål; 20. en nisse, en trykknap.

Training list T1: "Tine tænkte på ... og ... i går"

1. en flaske, en rosin; 2. en klud, en statue; 3. en gåtur, en drage; 4. et forår, et punktum; 5. et halsbånd, en spids; 6. et album, en pose; 7. en gulerod, et skab; 8. en skurvogn, en hætte; 9. en skjorte, et skib; 10. en fabrik, en figur; 11. et kort, en trætop; 12. en salami, en sofa; 13. en række, en tunfisk; 14. en sportshal, et dæk; 15. en abrikos, en sti; 16. en skovsnegl, en bus; 17. en fjer, en kornmark; 18. en robåd, en skinke; 19. et ønske, et lyntog; 20. en tante, en sørøver.

Acknowledgments

This research was supported by a PhD stipend from the William Demant Foundation. The authors wish to thank Signe Hjorth Fogh and Nicoline Gotholt Madsen for their help with the data collection. The børneDAT material is available upon request to the corresponding author.

Declaration of Conflicting Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Shno Koiek https://orcid.org/0000-0001-5779-3599
Laila Kjærbæk https://orcid.org/0000-0003-1428-8220
Tobias Neher https://orcid.org/0000-0002-1107-9274

References

Dunn, L. M., Dunn, L. M., Bulheller, S., & Häcker, H. (1965). *Peabody picture vocabulary test*. American Guidance Service Circle Pines.

Elberling, C., Ludvigsen, C., & Lyregaard, P. E. (1989). DANTALE: A new Danish speech material. *Scandinavian Audiology*, *18*(3), 169–175. https://doi.org/10.3109/01050398909070742

Firmansyah, D. (2018). Analysis of language skills in primary school children (study development of child psychology of language). *PrimaryEdu-Journal of Primary Education*, *2*(1), 35–44. https://doi.org/10.22460/pej.v1i1.668

Johnson, C. E. (2000). Childrens' phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, *43*(1), 144–157. https://doi.org/10.1044/jslhr.4301.144

Moore, D. R., Whiston, H., Lough, M., Marsden, A., Dillon, H., Munro, K. J., & Stone, M. A. (2019). FreeHear: A new sound-field speech-in-babble hearing assessment tool. *Trends in Hearing*, *23*, 2331216519872378. https://doi.org/10.1177/2331216519872378

Neumann, K., Baumeister, N., Baumann, U., Sick, U., Euler, H. A., & Weißgerber, T. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *International Journal of Audiology*, *51*(3), 157–163. https://doi.org/10.3109/14992027.2011.633935

Nielsen, J. B., & Dau, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology*, *48*(10), 729–741. https://doi.org/10.1080/14992020903019312

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International Journal of Audiology*, *50*(3), 202–208. https://doi.org/10.3109/14992027.2010.524254

Nielsen, J. B., Dau, T., & Neher, T. (2014). A Danish open-set speech corpus for competing-speech studies. *Journal of Acoustical Society of America*, *135*(1), 407–420. https://doi.org/10.1121/1.4835935

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Shield, B. M., & Dockrell, J. E. (2003). The effects of noise on children at school: A review. *Building Acoustics*, *10*(2), 97–116. https://doi.org/10.1260/135101003768965960

Soli, S. D., & Sullivan, J. A. (1997). Factors affecting children's speech communication in classrooms. *Journal of the Acoustical Society of America*, *101*, S3070. https://doi.org/10.1121/1.418755

Wagener, K., Brand, T., & Kollmeier, B. (2006). Evaluation des Oldenburger Kinder-Reimtests in Ruhe und im Störgeräusch [Evaluation of the Oldenburg children's rhyme test in silence and in noise]. *HNO*, *54*(3), 171–178. https://doi.org/10.1007/s00106-005-1304-4

Wagener, K., Josvassen, J. L., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology*, *42*(1), 10–17. https://doi.org/10.3109/14992020309056080

Walker, E., Sapp, C., Oleson, J., & McCreery, R. W. (2019). Longitudinal speech recognition in noise in children: Effects of hearing status and vocabulary. *Frontiers in Psychology*, *10*, 2421. https://doi.org/10.3389/fpsyg.2019.02421

Werner, L. A., & Boike, K. (2001). Infants' sensitivity to broadband noise. *The Journal of the Acoustical Society of America*, *109*(5), 2103–2111. https://doi.org/10.1121/1.1365112