The Danish hearing in noise test

Nielsen, Jens Bo; Dau, Torsten

Published in:
International Journal of Audiology

Link to article, DOI:
10.3109/14992027.2010.524254

Publication date:
2010

Document Version
Early version, also known as pre-print

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

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.
The Danish hearing in noise test

JENS BO NIELSEN & TORSTEN DAU

Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

Abstract

Objective: A Danish version of the hearing in noise test (HINT) has been developed and evaluated in normal-hearing (NH) and hearing-impaired (HI) listeners. The speech material originated from Nielsen & Dau (2009) where a sentence-based intelligibility equalization method was presented. Design: In the present study, the speech material was evaluated for naturalness and a subset of sentences selected. The new sentence lists were validated, and after three weeks retested. An additional experiment investigated how recollection of sentences affected the listeners’ performance. Study sample: 16 NH and 16 HI listeners participated in the validation and retest. Twelve HI listeners participated in the experiment on recollection. Results: The average speech recognition threshold in noise (SRT_N) for the NH listeners was -2.52 dB, with an overall standard deviation of 0.87 dB. The within-subject standard deviation was similar for the NH and the HI listeners. In the retest, the SRT_N decreased by 0.4 dB in both groups. Conclusion: The Danish HINT consists of 10 test lists and three practice lists each containing 20 sentences. The validation results are comparable to those of other versions of HINT. The test seems equally reliable for NH and HI listeners. After three weeks, reliable results can be obtained when sentence lists are reused with the same listeners.

Key Words: Speech intelligibility; Speech perception; HINT; Danish

Nielsen & Dau (2009) developed a speech intelligibility test in Danish, the conversational language understanding evaluation (CLUE), based on the principles of the original hearing in noise test (HINT; Nilsson et al, 1994). The CLUE test consists of 18 phonetically balanced test lists and seven practice lists. Each list contains 10 sentences. After its completion, the CLUE test was presented to the Danish hearing aid manufacturers Oticon, GN Resound, and Widex, and one of the companies conducted an extensive internal evaluation of the test. The evaluation acknowledged that the validation results for CLUE were comparable to those of the original HINT and to those of other language versions of the test, e.g. the Canadian-French version (Vaillancourt et al, 2005), the Cantonese version (Wong & Soli, 2005), and the Swedish version (Hällgren et al, 2006). However, the evaluation also indicated some concerns regarding (1) the speech material, (2) the choice of talker, and (3) the scoring rules, as outlined in the following.

1) The CLUE sentences were based on written materials like newspapers and magazines and fulfilled a set of criteria (Nielsen & Dau, 2009), but these criteria did not include an explicit requirement for simplicity in wording or contents. In contrast, typical HINT sentences are based on text materials that can be understood by 6–7 year-old children. The evaluation considered several sentences in the CLUE material as being unnatural or having a higher level of abstraction than typical for HINT sentences. Some sentences have inversion (reversed word order) and some verbs are in passive form. Furthermore, the evaluation considered some words and expressions to be ‘old-fashioned’.

2) The evaluation pointed out that the talker’s voice quality varies over time and that his pronunciation is ‘remarkable’ partly because of tension. The pronunciation of some of the sentences was considered less clear than that of others and the speed of
speech was perceived as slightly varying. It was argued that a more trained talker would be preferable in order to achieve a speech material as consistent and ‘transparent’ as possible.

3) The scoring rules for a sentence test determine when a listener’s response is considered correct, and these rules typically permit minor response deviations from the actual sentence. The CLUE scoring rules permit both some general variations, e.g. a change of verb tense, and a few specific variations. In the evaluation, it was argued that the CLUE scoring rules might cause less consistent scoring than would be desirable, and a clarification of the rules was recommended.

A project was established with the objective of creating a new speech intelligibility test that was based on CLUE but took the above-mentioned concerns into consideration. The test was to be validated for both normal-hearing (NH) and hearing-impaired (HI) listeners, since homogeneous test results for NH listeners do not necessarily imply homogeneous results for HI listeners (McArdle & Wilson, 2006). In addition, the goal was to create a test that corresponds to the current HINT standard (Bio-logic Systems Corp., 2005), such that it can be referenced as ‘the Danish HINT’. The standard demands test lists with 20 sentences per list, which is twice as many as in the original HINT and in CLUE. It was assumed that an improved speech material for the new test could be achieved by exchanging some of the CLUE test sentences with sentences from the CLUE practice lists.

As part of the project, effects of learning, typically involved in a sentence test, were investigated. Here, learning is considered as the combination of two separate effects; one related to ‘practice’ and one related to ‘memory’. The practice effect is associated with performance improvements that follow from getting more experience with the test; the memory effect is associated with performance improvements that follow from the recollection of specific sentences. The practice effect evolves continuously with each sentence that is presented, while memory only affects performance when sentences are reused with the same listener. In the validation experiments of the present study, a practice effect was estimated from the improvements in the listeners’ performance during their first test session. In a validation retest after three weeks, where the sentence lists from the first test session were reused, a combined learning effect (practice and memory) was assessed. The distribution of the learning effect between practice and memory in such a retest was investigated in an additional experiment with HI listeners.

**From CLUE to a Danish HINT**

**Test of naturalness**
The naturalness of the CLUE sentences was judged by a panel of 10 native and ‘naive’ Danish speakers and by two professional linguists. For various reasons, 15 of the practice sentences were rejected in advance, leaving 235 sentences for the naturalness test. The panel judged the written version of the sentences on a scale from 1 (= ‘artificial’) to 7 (= ‘natural’). The requirements for a sentence to be ‘natural’ were (1) that it did not contain unusual Danish words; and (2) that it could have been used in an ordinary conversation. A mean rating of 5 or above among the naive participants was set as the requirement for including a sentence in the test lists. In addition, up to three sentences with a score between 4.0 and 4.9 would be accepted in each test list. A score of 5 or above was achieved by 176 sentences, and 41 sentences received a score between 4.0 and 4.9. A sufficient number of ‘natural’ sentences were thus available to compile 10 new 20-sentence lists.

**Generation of the test lists**
The 18 original CLUE test lists and two of the CLUE practice lists were combined to create ten 20-sentence lists. The CLUE list with the lowest mean speech recognition threshold in noise (SRT_N), as determined during the CLUE validation process, was successively paired with the list with the highest SRT_N in an attempt to achieve lists with equalized SRT_N. In these lists, the ‘unnatural’ sentences were exchanged with sentences from the pool of ‘natural’ sentences, preferring those with a higher naturalness score. The exchanged sentences were reshuffled among the lists by a computer-based trial-and-error routine in order to maintain the phonetic balance between the lists as closely as possible (Nielsen & Dau, 2009). It was observed that the 24 sentences with a naturalness score of 4.0 to 4.9 had been distributed with two or three sentences in each list. Three practice lists were compiled from the sentences that were deemed ‘unnatural’ or omitted at previous stages.

**Permitted response variations**
A new set of rules for permitted variations in the listener response was created for the Danish HINT; the main difference from the CLUE scoring rules was the omission of alternatives for some specific words. The final scoring rules for the Danish HINT permit the following response variations: (1) change in verb tense; (2) change in article; (3) change between singular and plural nouns; (4) reordering of words; (5) addition of extra words or phones; and (6) omission of a single phone (e.g. the [t] that changes adjectives to adverbs in Danish). Several variations are permitted in a single response.

**Test validation with NH and HI listeners**
The purpose of the validation was to establish normative data for the test and to investigate the test reliability. Normative data can only be established for NH listeners, whereas the reliability can be judged for both NH and HI listeners from the within-subject standard deviation of the SRT_N and the variation of the mean SRT_N of the test lists. All listeners participated in a retest after three weeks.

**Method**

**LISTENERS**
Sixteen (8 male, 8 female) NH listeners and sixteen HI listeners (10 male, 6 female) participated in the validation. Participation was approved by the ethics committee of Copenhagen County. The NH
listeners’ age was between 19 and 43, with a mean of 33.6 years. The requirements for participation were: (1) age 18–45 years; (2) hearing threshold ≤ 20 dB HL at both ears (0.125 to 8 kHz), yet a threshold of 25 dB HL was allowed at one frequency per ear; (3) Danish as native language; (4) no previous experience with CLUE; and (5) variation in the educational background for the group.

The age of the HI listeners was between 61 and 69 (mean 65.9 years) and the requirements for participation were: (1) Age 60–70 years; (2) a hearing loss caused by presbyacusis, reflecting symmetrical mild-to-moderate sloping hearing loss; (3) at least one year of experience with wearing a hearing aid; (4) Danish as native language; (5) experience with DANTALE II (Wagener et al, 2003); (6) no previous experience with CLUE; and (7) variation in the educational background for the group.

**APPARATUS AND PROCEDURE**

The validation tests took place in a soundproof booth and the stimuli were presented diotically over Sennheiser HD580 headphones. The sound level was calibrated using the ear simulator and flat plate adaptor specified in IEC 60318-1 (2009), and a Bruel and Kjær measuring amplifier (type 2636). All testing was conducted without the use of hearing aids. The tests were conducted according to the standard HINT procedure (Bio-logic Systems Corp., 2005), controlled by a MATLAB application. The order of the sentences within each list was randomized before presentation of the list. The listeners received oral instructions before the test and were encouraged to guess if necessary when responding to the presented sentences. Each listener was tested with all 10 test lists. The order of the test lists was counterbalanced across listeners (using Latin squares) to avoid order effects. A short break was included after completion of the first five lists.

In order to familiarize the listeners with the task and to reduce the practice effect during the validation, a training procedure was conducted before the actual test. For the NH listeners, two practice lists in noise were presented. For the HI listeners, this procedure was preceded and extended by two practice lists in quiet in order to introduce the test smoothly and to determine an appropriate noise level for the subsequent list presentations in noise. The speech recognition threshold in quiet (SRT\textsubscript{Q}) of the second practice list determined the level of the noise. If SRT\textsubscript{Q} ≤ 45 dB(A), the noise level was fixed at 65 dB(A). If SRT\textsubscript{Q} > 45 dB(A), the noise level was fixed at SRT\textsubscript{Q} + 20 dB. This determination of the level for HI listeners followed the current HINT recommendations (Bio-logic Systems Corp., 2005). For the NH listeners, the noise level was always fixed at 65 dB(A).

The retest three weeks later followed the same schedule and procedure as the test, except that the practice lists in quiet were not presented. The individual noise levels determined during the first visit were also used in the retest. The order of the lists was the same as during the test, but the randomization of the sentences within the lists was different.

**Results**

**VALIDATION**

All SRT\textsubscript{N}s in the present study were calculated according to the current HINT standard (Soli & Wong, 2008). The overall SRT\textsubscript{N} across test lists and NH listeners was −2.52 dB with a standard deviation of 0.87 dB; the within-subject standard deviation was 0.86 dB. For the HI listeners, the overall SRT\textsubscript{N} was 0.09 dB with a standard deviation of 1.79 dB; the within-subject standard deviation was 0.92 dB.

For each of the 10 lists, a mean list-SRT\textsubscript{N} across the listeners was calculated. A normalized result is shown in Figure 1 for the NH listeners (black circles) and the HI listeners (grey circles). For the NH listeners, the list-SRT\textsubscript{N} standard deviation was 0.32 dB and the maximum deviation from the overall mean was 0.63 dB. For the HI listeners, the list-SRT\textsubscript{N} standard deviation was 0.39 dB and the maximum deviation from the overall mean was 0.60 dB. The normalized list-SRT\textsubscript{N}s were similar for the two groups; the largest deviation of 0.50 dB was observed for list 2. However, even for this list, an unpaired t-test did not show a significant difference between the list-SRT\textsubscript{N}s for the two groups [p = 0.15].

For the NH listeners, a two-way ANOVA showed a significant effect of list at a 0.05 level but not at a 0.01 level [F (9, 135) = 2.37, p = 0.016]. There was no significant effect of listener [F (15, 135) = 1.34, p = 0.19]. A corresponding analysis of the HI data showed a significant effect of list [F (9, 135) = 3.28, p = 0.0012], and a highly significant effect of listener [F (15, 135) = 35.31, p < 0.0001].

Figure 2 shows the mean SRT\textsubscript{N} across the 10 test lists for each of the NH listeners (black circles) and each of the HI listeners (grey circles). The subject-SRT\textsubscript{N} variation among the HI listeners was 6 dB and thus much larger than for the NH listeners (1.1 dB).

**PSYCHOMETRIC FUNCTIONS**

The psychometric function of the test was determined for each individual listener. The data points were based on the percentage of correctly repeated sentences at each of the signal-to-noise ratios (SNRs) of the adaptive procedure. (The adaptive procedure makes presentations that only deviate 0.2 dB SNR from each other possible. These presentation levels were pooled in bins of one dB around the integer values of the SNR.) The sentences at list positions 5–20 in the 10 test lists were included in the calculation, resulting in 160 data points for each listener. For each listener, a cumulative normal distribution function was fitted to the data, estimating a psychometric function. For the NH listeners, the steepest slope of these curves varied from 10.9 to 20.7 %/dB with a mean value of 16.8 %/dB. For the HI listeners, the steepest slope varied from 7.5 to 24.1 %/dB with a mean value of 14.7 %/dB. The steepest slopes of the psychometric functions are shown in Figure 3 as a function of the corresponding subject-SRT\textsubscript{N} for each listener. For the HI listeners (grey circles),
there was a significant correlation between the slope of the psychometric function and the $SRT_N$ $[r = -0.65]$. For the NH listeners (black circles), no significant correlation was found $[r = -0.03]$. An unpaired t-test did not show a significant difference between the mean of the steepest slopes for the 16 NH listeners (black square) and that for the 16 HI listeners (grey square) $[p = 0.15]$.

**Practice effect during the test**

Figure 4 shows the mean $SRT_N$ as a function of the list position during the test sessions. For each position, the $SRT_N$ was determined as the mean across the combinations of listeners and lists at that position during the test ($n = 16$), calculated separately for the NH listeners (black circles) and the HI listeners (grey circles). The data were normalized with respect to list-$SRT_N$ and subject-$SRT_N$, i.e. the effects of list and listener were removed. A linear regression line was fitted to the data for the 10 list positions; the slopes were (with 95% confidence intervals): $-0.05 [-0.09, -0.008]$ dB/position for the NH listeners and $-0.025 [-0.08, 0.03]$ dB/position for the HI listeners. For the NH listeners, the major effect of practice seemed to occur during the two first list presentations. If these two presentations were taken out of the linear regression, the slope would reduce to $-0.027 [-0.08, 0.03]$ dB/position. Thus, a significant practice effect was only observed for the NH listeners and only when the effect was considered over all 10 list presentations.

**Test-retest learning effect**

Figure 5 compares the list-$SRT_N$’s in the test (filled symbols) and the retest (open symbols). Accordingly, Figure 6 compares the results for the NH listeners (filled and open squares); the upper curves compare the results for the HI listeners (filled and open circles).
subject-SRTNs in the test (filled symbols) and in the retest (open symbols). In the retest, the overall SRTN across test lists and listeners was $-2.94 \text{ dB}$ for the NH listeners, a decrease of $0.42 \text{ dB}$ compared to the initial test due to learning effects (practice and memory). The overall SRTN standard deviation was $0.75 \text{ dB}$ and the within-subject standard deviation was $0.69 \text{ dB}$. A two-way ANOVA showed no significant effect of list [$F (9, 135) = 1.31, p = 0.24$], but a significant effect of listener [$F (15, 135) = 3.06, p = 0.0003$].

For the HI listeners, the overall SRTN in the retest was $-0.27 \text{ dB}$, a decrease of $0.36 \text{ dB}$ compared to the initial test. The overall SRTN standard deviation was $1.86 \text{ dB}$ and the within-subject standard deviation was $0.83 \text{ dB}$. A two-way ANOVA showed no significant effect of list [$F (9, 135) = 1.09, p = 0.37$], but a highly significant effect of listener [$F (15, 135) = 44.6, p < 0.0001$].

Discussion
The Danish HINT evaluated in this study produces normative data that are comparable to other language versions of HINT. The SRTN for the NH listeners is $-2.5 \text{ dB}$, which falls slightly outside the range of $-5.3$ to $-2.6 \text{ dB}$ observed for the 13 versions of HINT listed in Søli & Wong (2008). The relatively high SRTN for the Danish test might be caused by the complexity of the sentences and the use of a non-professional talker. This does not necessarily represent a disadvantage of the test. One of the goals of creating a new test was to achieve a normative SRTN that is considerably higher than that of existing Danish tests such as the DANTALE II test ($-8.4 \text{ dB}$; Wagener et al, 2003). The normative standard deviation of the SRTN for the Danish HINT, $0.87 \text{ dB}$, is similar to the mean for the HINTs reported in Søli & Wong (2008).

The observed effect of list in the validation test with NH listeners corresponds to the results obtained for the American HINT (Nilsson et al, 1994) and the Swedish HINT (Hällgren et al, 2006). Although the list effect is significant at a $0.05$ level, a post-hoc analysis with a Bonferroni correction ($n = 10$) showed that none of the list-SRTNs deviated significantly from the overall SRTN at a $0.05$ level. A similar result is obtained when performing a post-hoc analysis of the validation data for the HI listeners. Thus, the post-hoc analysis of the validation results does not indicate that certain lists should be avoided when using the Danish HINT for SRT measurements.

The overall SRTN for the HI listeners ($0.09 \text{ dB}$) was found to be $2.6 \text{ dB}$ higher than for the NH listeners ($-2.52 \text{ dB}$). This suggests that the test is sensitive to the listeners’ ability to follow a conversation in noise. For the HI listeners, the noise level was fixed at 20 dB above the SRTN (or minimum $65 \text{ dB(A)}$). This approach reduces the role of audibility and increases the sensitivity of the SRTN to other speech-reception difficulties such as cognitive factors. However, reduced audibility in some frequency regions may still explain part of the poorer performance for some of these listeners.

The within-subject standard deviation of $0.92 \text{ dB}$ for the HI listeners was found to be only marginally larger than the value of $0.86 \text{ dB}$ for the NH listeners. Thus, the reliability of the test seems similar for the two groups. However, this result may partly be explained by the HI listeners’ previous experience with DANTALE II; this was one of the requirements for their participation in the present study. Trained listeners are typically more focussed on the task and show a more reliable performance than untrained listeners. This may have reduced the within-subject standard deviation.

During the presentation of the 10 test lists, the practice effect was small in both listener groups, but particularly small for the HI listeners. If the two first test lists were omitted from the calculations for the NH listeners, the effect would reduce to the same level as for the HI listeners. This suggests that the smaller effect observed for the HI listeners could be due to the two additional practice lists that were presented before the actual test session. It thus seems that running four practice lists instead of only two can significantly reduce the progression of the practice effect during the following list presentations.

The similar results obtained in the test and the retest both for NH and HI listeners suggest that the test can be reused after three weeks. The decrease of the overall SRTN of $0.4 \text{ dB}$ from test to retest for both listener groups is too small to affect the functionality of the test. Furthermore, the within-subject standard deviation was reduced in the retest and the significant effect of list observed in the initial test was not observed in the retest.

Effects of practice and memory
An additional experiment was performed with a new group of HI listeners. The purpose was to estimate how the learning effect is distributed between practice and memory when sentences are reused with the same listeners. The effects were estimated from the difference in the listeners’ average performance during an initial test and a retest. The within-session progression of the practice effect, as depicted in Figure 4, was not investigated here.

Method

LISTENERS
Twelve (9 male, 3 female) HI listeners participated. Participation was approved by the ethics committee of Copenhagen County. Their age was between 59 and 72 years, mean 64.8 years. The requirements for the listeners in this group were the same as for the previous HI group (although the age requirement was slightly violated for three listeners).

PROCEDURE
The experiment was divided in two sessions; the second visit took place three weeks after the first (live and a half weeks later for one of the listeners). The practice and the test procedures were similar to...
those of the test validation experiments. The only major difference was that only five test lists were presented at the first visit. During the experiment, subsets of the 10 test lists were presented in three conditions: (1) five unknown lists presented at the first visit (‘first visit test’); the test results in this condition were not affected by any memory effect; (2) five unknown lists presented at the second visit (‘second visit test’); these results were affected by the progression of the average practice effect between the first and the second visit, but still not affected by memory; and (3) the five lists from the first visit presented again at the second visit (‘second visit retest’); the results in this condition were affected by both memory and a change in the average practice effect. The practice component of the learning effect can thus be estimated from the SRT\textsubscript{N} difference between ‘first visit test’ and ‘second visit test’. The memory component can be estimated from the SRT\textsubscript{N} difference between ‘second visit test’ and ‘second visit retest’.

The test lists of the first visit were counterbalanced across listeners and each list was included in half (six) of the subsets. The order of the lists was also counterbalanced to avoid order effects. During the second visit, the order of the five previously presented lists was the same as during the first visit. The five new lists and their order were counterbalanced across listeners. The previously presented and the new lists interleaved through the second session.

**Results**

Three mean SRT\textsubscript{N}s were calculated for each listener: (1) the mean SRT\textsubscript{N} across the five lists presented at the first visit; (2) the mean SRT\textsubscript{N} across the five lists presented for the first time at the second visit; and (3) the mean SRT\textsubscript{N} across the five lists presented for the second time during the second visit. For each listener, the means were normalized with respect to the mean SRT\textsubscript{N} of the ‘second visit test’ in order to remove the large SRT\textsubscript{N} differences between listeners. The results are shown in Figure 7. The mean SRT\textsubscript{N}s across listeners in the three conditions were: 0.10 dB for ‘first visit test’; 0 dB for ‘second visit test’; and −0.15 dB for ‘second visit retest’.

The estimate for the change in the average practice effect from test to retest is thereby −0.10 dB. The pure memory effect is estimated to be −0.15 dB.

**Discussion**

The difference in learning effect (practice and memory) from test to retest can be estimated to −0.25 dB. This is slightly lower than for the previous group of HI listeners (−0.36 dB), probably because only five test lists were presented during the first visit (instead of 10).

The memory effect (−0.15 dB) seems to be slightly larger than the practice effect (−0.10 dB). However, the estimate of the memory effect was dominated by the particularly large effect observed for listener 1. Omitting this result from the calculation would reduce the memory effect to −0.04 dB.

The results from the present experiment confirm the results obtained with the NH listeners and the previous group of HI listeners (Figure 6) that the SRT\textsubscript{N} change between test and retest is within 0.5 dB for most listeners. The results also indicate that only half or probably less of the SRT\textsubscript{N} decrease between test and retest is due to a memory effect.

**Conclusion**

A Danish HINT with 10 test lists and three practice lists was developed. The test lists and practice list are shown in the Appendix. The test validation with NH listeners produced normative data that are comparable to those of other language versions of HINT (Soli & Wong, 2008). The normative SRT\textsubscript{N} of −2.5 dB for the Danish HINT is slightly above that obtained with other HINTs, and it is substantially higher than the value obtained with another Danish speech test, DANTALE II (Wagener et al, 2003).

The validation with HI listeners led to SRT\textsubscript{N} assessments with a within-subject deviation and a between-list deviation that was only slightly different from those obtained with NH listeners. The test is thus expected to produce equally reliable results for NH and HI listeners.

The test and retest with a three-week interval showed only small differences in the measured SRT\textsubscript{N}s. Changes in the subject-SRT\textsubscript{N}s were generally within 0.5 dB. Reuse of the test lists after three weeks thus seems possible. The investigation of the separated practice and memory effects suggested that recollection of the sentences only accounts for a minor part of the SRT\textsubscript{N} decrease between test and retest.

**Acknowledgements**

This study was conducted in collaboration with the Danish hearing aid companies Oticon, GN Resound, and Widex. We wish to thank the following colleagues for their contributions and involvement: Lise Bruun Hansen and Niels Sogaard Jensen (Oticon); Anja Kofod Pedersen and Ellen Raben Pedersen (Widex); Charlotte T. Jespersen, Jenny Nøsgaard, and Lotte Hernvig (GN Resound). We would also like to thank Julie Neel Weile for running the tests with the many HI listeners. Finally, thank you to all the listeners who took time to participate. The present work was partly funded by the Oticon Foundation.

**Note**

Audio files of the Danish HINT and the written sentence lists can be found at http://informahealthcare.com/doi/abs/10.3109/1499202.7.2010.524254.
Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References


Supplementary material available online

Appendix
Appendix.

Test list 1
1) Det var en god fastelavnsfest
2) Kampen skal spilles på onsdag
3) Filmen er rigtig godt lavet
4) Derhjemme spiser vi ikke kød
5) Børnene løber rundt og leger
6) Hun kommer meget i teatret
7) Familien går tur i parken
8) Statuen har ikke noget hoved
9) Hun tog en hurtig beslutning
10) Vi snakked med vores venner
11) Billetterne bliver sendt til os
12) Bussen kan ikke komme frem
13) Posen her er til grøntsager
14) Han sluttede som nummer fire
15) Chokoladen var dyr og god
16) Byen ser fantastisk dejlig ud
17) Jeg skulle ringe til formanden
18) Vi sagde farvel til gæsterne
19) Bakken er halvtreds meter høj
20) Arbejdet er hårdt og krævende

Test list 2
1) Reden er bygget af smågrene
2) Jeg ønsker mig et kæledyr
3) Han var verdensmester i svømning
4) De cykler eller tager bilen
5) Huset lå omme bag torvet
6) Jeg spurgte ikke til prisen
7) De ankom sidst på formiddagen
8) Hun rider på en fællevirksomhed
9) Insekter kan flyve meget langt
10) De har altid boet hjemme
11) Mødet skal holdes på skolen
12) Udenfor er det fuldstændigt mørkt
13) Hun var omgivet af mennesker
14) Børnene kom hjem ved middagstid
15) Snakket ved bordet var livlæg
16) Alle foredrag er på engelsk
17) Af og til laver naboerne
18) De blev hurtigt gode venner
19) Han afviste det nye forslag
20) Han lagde tasken på bordet

Test list 3
1) Om morgenen lægde stormen sig
2) Lyden kommer oppe fra loftet
3) Hun har købt en vinterfrakke
4) Grisene løber frit på marken
5) Han talte til en kollega
6) Musik giver en god stemning
7) Spillerne troede på sig selv
8) Tapetet var faldet af væggen
9) Han havde de smukkest øjne
10) Hver aften spiser de salat
11) Mandag vågnede vi meget sent
12) Hendes far var ikke hjemme
13) Han er tilfældigt med artiklen
14) Klokken var blevet over midnat
15) Båndet blev revet i stykker
16) Butikken holder et stort udsalg
17) Hun lavede en kop kaffe
18) Nu venter landmandene på regn
19) De kommer sejlende til byen

Test list 4
1) Pigen strikker en rød trøje
2) Vi ventede længe i køen
3) Om aftenen var der lejrtilbæk
4) Det kilder lidt i fingeren
5) Hun gik hen til telefonen
6) Vi skal bare blive siddende
7) Kunden er tilfreds med svaret
8) Huset her er hans barndomshjem
9) Redskaber skal sættes på plads
10) Vejrudsigten lover regn og slud
11) Godt håndværk holder i år
12) Min kuglepen skriver med rød
13) Mødet sluttede efter tre timer
14) Han ønskede sig en jakt
15) Jeg er ikke længere sulten
16) Han købte ikke mange blomster
17) Villaen er ikke blevet solgt
18) Hjælpen nåede frem for sent
19) Hendes bror vil være brandmand
20) Han lagde tasken på bordet

Test list 5
1) Børnene sidder i en rundkreds
2) Gæsterne nyder den gode vin
3) Han var i oppidning om mennesker
4) Børnene kom hjem ved middagstid
5) Snakket ved bordet var livlig
6) Alle foredrag er på engelsk
7) Af og til laver naboerne
8) De blev hurtigt gode venner
9) Han afviste det nye forslag
10) Begge hold scorede otte mål
11) Stuen skal nok blive hyggelig
12) Døren er næsten aldrig åben
13) Han blev en god skolelærer
14) De engelske bøffer var mere
15) Han kunne kore meget stærkt
16) Sofanen står bagerst i rummet
17) Torsdag var han ikke hjemme
18) Begge fodboldhold klarer sig fint
19) Maden blev serveret til tidligere
20) Han havde let ved hovedregning

Test list 6
1) Nu skal maskinerne skifte ud
2) Renten var kun fire procent
3) Jeg tager fat i dørhåndtaget
4) Tøjet var gået af mode
5) Her går alle med solbriller
6) Kassedamen så venligt paa ham
7) Han ligger stadig i sengen
8) Eleven skriver en lang rapport
9) Hjem hentet kom til brylluppet
10) Vi så lidt af vejrudsigten
11) Tøget er meget sjældent fuldt
12) Jeg var også utrolig glad
13) Hans datter vil på højskole  
14) I går havde filmen premiere  
15) Fabrikkens port var ikke lukket  
16) Hendes tøj var helt gennemblødt  
17) Bilen er ikke længere ny  
18) Nu begynder en ny sæson  
19) Flyrejsen varer mindst fem timer  
20) Jeg sætter mig nede bagved

Test list 7
1) Lakken skal fjernes fra gulvet  
2) Han kan lugte hendes parfume  
3) Værelset lår ud til baggården  
4) Naboorne var med til middagen  
5) Lyskrydset skifter snart til rødt  
6) Han er en flettig musiker  
7) Vi havde en dejlig weekend  
8) Udsigten er bedst om sommeren  
9) Hendes øjnene så trætte ud  
10) Vi får boller og chokolade  
11) Skuffen kunne ikke lukkes helt  
12) Vi byggede husene af træer  
13) I morgen bliver vejret bedre  
14) Han hoppede op på cyklen  
15) Han har aldrig lavet middagsmad  
16) Udsigten til skoven var god  
17) Motorløb kan være ret farligt  
18) Vi rister pølser over bålet  
19) Manden kom til en benzintank  
20) Vi er en fredelig familie

Test list 8
1) Han læser med stærke briller  
2) Pludselig kom der en lastbil  
3) Der var altid åbent tirsdag  
4) Mine venner går i gymnasiet  
5) Bogen er skrevet på engelsk  
6) Der bor mange mennesker her  
7) Hun var taget på arbejde  
8) Vi ligner hinanden ret meget  
9) Vi skriver tillykke og skåler  
10) Chaufføren ser ind i spejlet  
11) Drys retten med hakket persille  
12) Drengen stikker hånden langt frem  
13) Han har passet sin træning  
14) Blomster og gaver strømmede ind  
15) Udsigten er bedst om sommeren  
20) Han dukser var meget korte

Practice list 1
1) Pigerne går rundt i haven  
2) Hendes ansigt er stadig solbrænt  
3) Filmen blev straks en succes  
4) Jeg kan godt lide jazzmusik  
5) Vi siger tillykke og skåler  
6) Drengene går rundt i haven  
7) Lad os bare køre igen  
8) Han har passet sin træning  
9) Denne dag er det dejligt at have en dejlig weekend  
20) Vi er en fredelig familie

Practice list 2
1) Drengen blev medlem af klubben  
2) Ikke langt væk ligger rådhuset  
3) Flyttmænd har tit ømme muskler  
4) Nu mangler vi blot tallerkner  
5) Bogen er billigt på udsalg  
6) Cykler kan lejes mange steder  
7) I spisestuen var lyset tændt  
8) I går kom svalerne hertil  
9) Jeg havde cyklet i solskin  
10) Skolekampen drikker et glas melk  
11) Suppen smagte godt af tomat  
12) Vi spadserede en tur sammen  
13) En ung pige kommer gående  
14) Snart fylder rapporten ti sider
15) Børnene og de voksne sover
16) En taxa kørte langsomt forbi
17) Kaninen sprang ud gennem hullet
18) Näste deltager var smedens søn
19) Lågen bag dem smækkede i
20) Under bogen ligger en tegning

Practice list 3

1) Han rensede skærmen for støv
2) Katten kom listende helt stille
3) Katten spind i hendes arme
4) Trøj en er syet af bomuld
5) Blomsterne vokser i små skåle
6) De to venner deler arbejdet
7) De sidder længe i tavshed
8) Store bølger slog mod stranden
9) Den gamle mand smilede stort
10) I regnbuen ses alle farver
11) De kørte direkte til skolen
12) Maden var rig på vitaminer
13) Konen er ældre end manden
14) Penge skal sættes i banken
15) Fødselsdagen er først på tirsdag
16) Postbudet har to små børnebørn
17) Det blev en pragtfuld ferie
18) Filmen var aldrig rigtig sjov
19) Jeg samler på gamle møbler
20) Om mandagen holder jeg fri