

# **New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities**

Jana Besser, Adriana Zekveld, Sophia E. Kramer, Jerker Rönnerberg and Joost M. Festen

**Linköping University Post Print**

N.B.: When citing this work, cite the original article.

Original Publication:

Jana Besser, Adriana Zekveld, Sophia E. Kramer, Jerker Rönnerberg and Joost M. Festen, New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities, 2012, Journal of Speech, Language and Hearing Research, (55), 1, 194-209.

[http://dx.doi.org/10.1044/1092-4388\(2011/11-0008\)](http://dx.doi.org/10.1044/1092-4388(2011/11-0008))

Copyright: American Speech-Language-Hearing Association

<http://www.asha.org/default.htm>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-74353>

**New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities**

Jana Besser<sup>1</sup>, Adriana A. Zekveld<sup>1,2,3</sup>, Sophia E. Kramer<sup>1</sup>, Jerker Rönnerberg<sup>2,3</sup>, Joost M.

Festen<sup>1</sup>

Author Affiliations

1 = ENT/Audiology Department & EMGO+ Institute for Health and Care Research, VU University Medical Center Amsterdam, The Netherlands

2 = Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Sweden

3 = Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

Corresponding Author

Jana Besser, MSc.

VU University medical center

Dept. of ENT/Audiology, EMGO+ Institute

P.O. Box 7057

1007 MB Amsterdam

The Netherlands

E-mail: [j.besser@vumc.nl](mailto:j.besser@vumc.nl)

Telephone: +31-20-444 09 02

Fax: +31-20-444 09 83

### Abstract

**Purpose:** In this research, the authors aimed to increase the analogy between Text Reception Threshold (TRT; Zekveld, George, Kramer, Goverts, & Houtgast, 2007) and Speech Reception Threshold (SRT; Plomp & Mimpen, 1979) and to examine the TRT's value in estimating cognitive abilities that are important for speech comprehension in noise.

**Method:** The authors administered 5 TRT versions, SRT tests in stationary (SRT<sub>STAT</sub>) and modulated (SRT<sub>MOD</sub>) noise, and 2 cognitive tests: a reading span (RSpan) test for working memory capacity and a letter–digit substitution test for information-processing speed. Fifty-five adults with normal hearing (18–78 years, M = 44 years) participated. The authors examined mutual associations of the tests and their predictive value for the SRTs with correlation and linear regression analyses.

**Results:** SRTs and TRTs were well associated, also when controlling for age. Correlations for the SRT<sub>STAT</sub> were generally lower than for the SRT<sub>MOD</sub>. The cognitive tests were correlated to the SRTs only when age was not controlled for. Age and the TRTs were the only significant predictors of SRT<sub>MOD</sub>. SRT<sub>STAT</sub> was predicted by level of education and some of the TRT versions.

**Conclusions:** TRTs and SRTs are robustly associated, nearly independent of age. The association between SRTs and RSpan is largely age dependent. The TRT test and the RSpan test measure different nonauditory components of linguistic processing relevant for speech perception in noise.

**Key Words:** masked text recognition, speech perception, cognitive abilities, age, working memory

## Introduction

A key component of human communication is the exchange of spoken information. Distortion of conversational interaction by factors such as hearing loss can have serious implications for the affected person's social daily life functioning (Nachtegaal et al., 2009; Stephens & Kramer, 2010). Consequently, it is of great interest to study the mechanisms underlying speech comprehension. Knowledge of these mechanisms is essential when considering strategies of rehabilitation – and, thus, communicative re-enablement – tailored to the individual challenges of a person with hearing impairment (Gatehouse, Naylor, & Elberling, 2003). Although research in the field of speech comprehension initially focused mainly on the peripheral auditory domain (Humes, 1991; Plomp, 1986), recent studies underline the importance of cognitive factors for speech understanding in adverse conditions (e.g., Akeroyd, 2008; Houtgast & Festen, 2008; Wingfield & Tun, 2007). Consequently, attention is also being directed to the influence of nonauditory parameters on speech comprehension, which has resulted in the research field of cognitive hearing science (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009).

During speech comprehension, bottom-up processing of acoustic signals in the auditory system interacts strongly with top-down cognitive mechanisms of stimulus interpretation in a single limited-capacity processing system (Wingfield & Tun, 2007). The involvement of cognition becomes more pronounced when listening conditions are challenging because of background noise or hearing impairment (Pichora-Fuller, Schneider, & Daneman, 1995; Schneider, Daneman, & Pichora-Fuller, 2002; Stenfelt & Rönnberg, 2009). These findings were implemented in the ease of language understanding (ELU) model (Rönnberg, 2003; Rönnberg, Rudner, Foo, & Lunner, 2008). This model assumes that explicit and effortful cognitive processes become more relevant if implicit matching of the linguistic input to stored phonological representations in long-term memory is not sufficient for lexical

access and signal interpretation. Reliable and feasible tests of the cognitive abilities demanded in such situations would be of great value for clinical purposes. Such cognitive tests could indicate whether a purely auditory rehabilitation program is likely to restore a person's communicative abilities. In the case of observed cognitive declines, a different rehabilitation strategy might be considered.

In an attempt to cover the relevant cognitive aspects of speech comprehension in one single test, which is also feasible and can be considered as a candidate for use in the clinic, the Text Reception Threshold (TRT) test was developed (Zekveld, George, Kramer, Goverts, & Houtgast, 2007). It was intended to function as a visual equivalent to the Speech Reception Threshold (SRT) test (Plomp & Mimpen, 1979) and builds on the assumption that the processing resources governing speech and text understanding overlap once the perceptual input stage has been passed (Humes, Burk, Coughlin, Busey, & Strausner, 2007). In both the SRT and the TRT tests, sentences are presented that are masked auditorily or visually according to the test modality. The participant's task is to repeat the presented sentences as accurately as possible. In an adaptive procedure, the degree of masking is manipulated based on the correctness of the participant's responses. The SRT and TRT tests were designed to be equivalent regarding level of difficulty and response complexity. Access to the stimulus information is limited in both tests because of the masker. Thus, the ability to make use of linguistic context is an inherent requirement in the tasks because the informational gaps need to be filled in by utilizing the available linguistic information. The integration and filling in of linguistic context information is a relevant ability for successful speech comprehension (Bronkhorst, Bosman, & Smoorenburg, 1993; Dubno, Ahlstrom, & Horwitz, 2000; Pichora-Fuller, 2008; Wingfield, 1996). This is especially important for people with hearing impairment, considering that they have to cope with fragmentary verbal information relatively often due to their auditory processing deficits (Schneider et al., 2002).

Zekveld et al. (2007) found that the TRT is significantly associated with the SRT in stationary noise and in 16-Hz block-modulated noise, both with the spectral shape of the long-term average spectrum of the target speech. They found about 30% shared variance between SRT and TRT in a sample of individuals with normal hearing. According to the authors, this proportion can be interpreted as the degree to which interindividual, modality-independent differences in cognitive and linguistic abilities account for individual differences in masked text and speech comprehension. The association of the TRT with effortful speech perception in noise was furthermore evidenced by George et al. (2007) and Zekveld, Kramer, and Festen (2011). However, the analogy of the TRT and SRT tests can still be improved. An investigation by Kramer, Zekveld, and Houtgast (2009) indicated that the relation among the TRT, verbal working memory (WM) capacity, processing speed, and speech reception should be established more clearly. Possibly, the association of the TRT with speech understanding in noise can be strengthened by making the TRT rely more heavily on WM and processing speed, considering these factors have been identified as essential determinants in the higher level processing of speech by earlier research (e.g., Daneman & Carpenter, 1983). A review article by Akeroyd (2008), in which the author compiled a large number of studies on the relationship between speech understanding in noise and aspects of cognition, pointed out WM capacity as the most evident cognitive factor involved in speech comprehension. Also, in other studies, researchers stressed the central role of WM capacity for speech comprehension in adverse conditions (e.g., Foo, Rudner, Rönnberg, & Lunner, 2007; Lunner, 2003; Pichora-Fuller & Singh, 2006; Rudner, Foo, Rönnberg, & Lunner, 2009; Rudner, Foo, Sundewall-Thorén, Lunner, & Rönnberg, 2008). WM capacity is usually assessed by means of memory span tests (Bopp & Verhaeghen, 2005). The first linguistic WM span tests (i.e., reading and listening span) were developed by Daneman and Carpenter (1980). During the span test, the participant is presented with short lists of sentences. At the end of the list, the participant is

asked to recall the last word of each of the presented sentences. The number of the correctly recalled items determines the person's WM span. These early tests stressed the storage component of WM. However, WM is principally understood as a system not only for temporary information storage but also for real-time processing of information in complex cognitive tasks (Baddeley, 2000, 2003; Rönnerberg et al., 2008). Accordingly, memory span tests have been adapted to make stronger claims on the processing component of WM, which was not inherent in the original test. The central feature of these later tests is that they include a dual task, which requires not only storage but also manipulation of the incoming information. Thus, they are specifically designed to reflect the theoretical assumptions about the WM system (Shelton, Elliott, Hill, Calamia, & Gouvier, 2009). For instance, in addition to recalling sentence words, the newer tests require participants to judge the semantic correctness of every sentence directly after sentence presentation (Baddeley, Logie, Nimmosmith, & Brereton, 1985; Lunner, 2003; Rönnerberg, 1990; Rönnerberg, Arlinger, Lyxell, & Kinnefors, 1989).

The use of complex span tasks in hearing research is also supported by the finding that tasks requiring semantic processing of the input are better able to assess the cognitive contributions to speech comprehension (Pichora-Fuller & Singh, 2006). To our knowledge, a Dutch WM test fulfilling these requirements has not been released yet. Therefore, for this study, we created a new Dutch reading span test with a dual processing task, resembling the test described by Andersson, Lyxell, Rönnerberg, & Spens (2001). This test was previously used in combination with tasks of speech understanding in the Swedish and Danish languages, providing a fair amount of reference material for comparisons of our results to earlier findings (Foo et al., 2007; Hällgren, Larsby, Lyxell, & Arlinger, 2001; Lunner, 2003; Rudner et al., 2008).

Besides WM capacity, processing speed has been identified as a relevant top-down determinant in speech comprehension (Janse, 2009; Wingfield, 1996). Consequently, performance on the TRT test should also depend on an individual's cognitive processing speed. According to Jolles, Houx, Van Boxtel, and Ponds (1995), so-called substitution tests are commonly used for the assessment of general information-processing speed. These tests activate several cognitive processes simultaneously and do not make inferences about the specific processes involved. In contrast to tests such as lexical or semantic access (Rönnerberg, 1990), they are not domain specific and thus are suitable instruments when the relevant processes are not specified. A widely known test in neuropsychological research is the Digit-Symbol Substitution Test (Wechsler, 1955, 1981), which requires the participant to match certain symbols to digits within a short period of time. From this test, the Letter-Digit Substitution Test (LDST; Jolles et al., 1995) was derived. The LDST is less dependent on memory and complex visual processes than are the earlier substitution tests and, therefore, can be considered a more specific measure of information-processing speed (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006). Van der Elst et al. also released normative LDST data for healthy adults. We used the LDST to assess processing speed in the current study.

The main aim of this study was to revise the existing TRT test in such a way that it would function as a real visual equivalent of the SRT test. We aimed to implement the suggestions by Kramer et al. (2009) mentioned above. Kramer and colleagues performed an investigation of how parallel the TRT and the SRT actually were in terms of their associations with information-processing speed, age, and WM capacity. The authors failed to find a significant correlation between WM capacity and the TRT. Hence, they concluded that in order for the TRT to relate more closely to the SRT test, it should rely more heavily on all of these factors – most importantly, WM capacity and processing speed. Kramer et al. argued that the absence of a correlation between TRT and WM could be due to the use of a spatial

WM task rather than a verbal one. A verbal WM test should be more suitable for examining the role of WM capacity in TRTs and SRTs, considering these are also verbal tests that are processed in the phonological rather than the visuospatial system of WM – that is, according to the WM model described by Baddeley (2000). The current study is the first to investigate the associations of the TRT with verbal WM capacity – and, moreover, the first to relate them directly to individual performances in speech perception in noise.

We developed several new versions of the TRT test. Each of these modifies the original TRT in one or several aspects, which we assumed would strengthen associations with processing speed and WM capacity. For instance, we adapted rates and manner of stimulus presentation and the number of sentences to be remembered before giving a response. We expected that these modifications would also increase the TRT test's association with age and, most important, speech comprehension in noise.

To test our hypotheses, we conducted the original and the new TRT test versions together with SRT tests in stationary and modulated noise, the LDST for processing speed, and the reading span test for WM capacity. The use of stationary and modulated noise for the SRT was motivated by earlier research that found significant differences between the two noise conditions in their association with the TRT (George et al., 2007) and other cognitive measures (Lunner & Sundewall-Thorén, 2007; Rönnberg, Rudner, Lunner, & Zekveld, 2010). We consider the nonstationary masker to be more ecologically valid because it is more representative of everyday listening situations (Kramer, Kapteyn, Festen, & Tobi, 1996) but use the stationary masker for comparison. On the basis of previous studies (George et al., 2007; Lunner & Sundewall-Thorén, 2007; Rönnberg et al., 2010), we expected that the relation between SRT and TRT would be stronger for the modulated noise masker.

## **Method**

## Participants

The study included 56 adult participants ages 18–78 years with normal hearing for pure tones (0.5 kHz–4.0 kHz). The data of one participant was later excluded due to strongly deviating results (see Results section). The remaining 55 participants (40 female, 15 male) were evenly distributed over the age range ( $M = 44$  years,  $SD = 14.2$  years). A comparable level of education across ages was pursued. We distinguished seven levels of education, which we coded as 1 (primary), 2 (lower vocational), 3 (pre-vocational general secondary), 4 (secondary vocational), 5 (pre-university and senior general secondary), 6 (higher professional), and 7 (university). See Table 1 for further demographic details.

All participants were native Dutch speakers and had no diagnosis of dyslexia. They had normal or corrected- to-normal vision. Participants also needed to pass tests for color blindness (Ishihara, 1989) and near vision screening to be admitted to the study. For the near vision test, we created Dutch word charts that resemble the charts described in Bailey and Lovie (1980). For inclusion in the current study, participants had to be able to read words down to a size of 16 points at a distance of about 50 cm from the computer screen.

None of the participants reported a history of ear disease. Prior to the study, their hearing was examined with a pure-tone audiogram at octave frequencies from 500 Hz to 4000 Hz (see Table 1). People were excluded from participation if their hearing sensitivity was worse than 20 dB HL at any of the frequencies in the test ear (i.e., the better ear). For the nontest ear, thresholds of 30 dB HL at all frequencies were permitted with an allowance for a one-frequency dip up to 40 dB HL. Commonly pure-tone averages (PTAs) over the tested frequencies are used as an inclusion criterion. PTAs up to 20 dB HL–25 dB HL are considered normal hearing, and losses up to 40 dB HL indicate a mild hearing loss (British Society of Audiology, 2004; World Health Organization, 2011). In that sense, the selection

criteria for this study were stricter than common practice. We considered that good hearing in both ears would prevent any potential disturbances by a decreased overall hearing sensitivity.

## Tests

**SRTs.** We assessed participants' ability to understand speech in noise by measuring the SRT in noise. The participants were asked to literally repeat auditory sentences in background noise presented monaurally through headphones to their better ear. With an adaptive masking procedure, we determined the SRT – that is, the signal-to-noise ratio (SNR) at which a person was able to repeat 50% of the sentences in a test run. An SRT test run included 13 short ordinary Dutch sentences (Versfeld, Daalder, Festen, & Houtgast, 2000). A lower SNR indicates a better test result.

We measured SRTs with a so-called 1-up, 1-down adaptive masking procedure: When a person was able to reproduce the entire sentence correctly, the SNR was decreased for the next sentence; otherwise, it was increased by 2 dB (Plomp & Mimpen, 1979). The 1-up, 1-down manner resulted in an intelligibility level at which, on average, half the sentences could be perceived correctly. The initial SNR at the test onset was  $-7$  dB for the SRT tests in stationary noise ( $SRT_{STAT}$ ) and  $-12$  dB for the SRT tests in modulated noise ( $SRT_{MOD}$ ). The first sentence of a test run was repeated with a 4-dB decrement in the level of masking – that is, a 4-dB higher SNR – per repetition until the participant was able to reproduce the sentence literally. All subsequent sentences were presented only once. When the preceding sentence had been repeated correctly, the SNR of the following sentence was decreased by 2 dB. Otherwise, it was increased using the same step size of 2 dB. Thus, the SNR of every sentence differed by 2 dB from the SNR of the sentence preceding it. We conducted the test maintaining an overall sound level (signal plus noise) of 65 dB SPL. Accordingly, we digitally adapted the levels of both noise and speech in order to manipulate the SNR.

The *outcome* SNR was the average SNR of Sentences 5–14 from each list. Sentence 14 was not actually part of the test list, but its hypothetical SNR could be deduced from the correctness of the participant's reply to Sentence 13. We excluded the data of Sentences 1–4 from the calculation of the threshold in order to allow for initial processes of stabilization during the adaptive procedure of approaching an individual's threshold.

We administered SRT tests with two different noise maskers – that is,  $SRT_{STAT}$  and  $SRT_{MOD}$ . The overall spectral shape of both noises equaled the long-term average spectrum of the target speech. The modulated noise consisted of two frequency bands (below and above 1 kHz), each modulated with the amplitude variations from a random speech sample of the target talker in the corresponding bands (Festen & Plomp, 1990).

**TRTs.** For the assessment of the participants' ability to read masked text, we determined the TRT (Zekveld et al., 2007). During this test, sentences that were partly masked by a vertical bar pattern were displayed on a computer screen at different masking levels (see Figure 1). Participants read aloud the sentences as accurately as possible. The outcome of the test indicates the percentage of unmasked text, at which a participant was able to read 50% of the presented sentences correctly. As for the SRT test, a lower test score means a better test performance.

In each test, a list of 13 ordinary Dutch sentences was presented, derived from the same corpus as the SRT stimuli (Versfeld et al., 2000). As with the SRT test, the percentage of masking was adapted for every sentence in a 1-up, 1-down test procedure based on the correctness of the participant's response to the previous sentence. The first sentence was initially presented at a level of 40% unmasked text. It was then repeated with a decrease in masking of 12% for every repetition until the participant was able to read the complete sentence correctly. All subsequent sentences were presented only once. The change in masking for each of these remaining sentences was 6% with reference to the sentence

preceding it: When the sentence had been reproduced literally, the percentage of masked text in the following sentence was 6% higher than in the previous one. Otherwise, it was 6% lower. Zekveld et al. (2007) quantified this step size to make the testing procedure as parallel as possible to the SRT test.

The *outcome* TRT was the mean percentage unmasked text of Sentences 5–14 (see the scoring described earlier in the SRTs subsection). The SRT and TRT tests were similar in terms of test material, presentation rates, adaptive (de)masking, and scoring procedures.

In this study, we created four new TRT tests, each of which was a different modification of the original TRT test. The modifications were intended to make the test more demanding with regard to online processing speed and WM capacity (Kramer et al., 2009). The following subsections contain descriptions of all applied test versions.

*TRT<sub>ORIGINAL</sub>*. This was the unmodified original test developed and described by Zekveld et al. (2007). The participants read one sentence at a time. Sentences were being built up word by word with every word remaining in its original spatial position when subsequent words were added. The time interval that passed from the onset of the presentation of one word until the onset of the following word matched the duration of that word in the corresponding audio recording of the sentence (Versfeld et al., 2000). After the presentation of the last word, the complete sentence remained on the screen for 3,500 ms before it disappeared. Participants were allowed to start their response before the presentation of the sentence was completed.

*TRT<sub>500</sub>*. The test design was the same as that for the *TRT<sub>ORIGINAL</sub>*, except that the complete sentence remained on the screen for 500 ms instead of 3,500 ms. The test was designed to put higher demands on processing speed (see also Zekveld, Kramer, & Festen, 2011).

*TRT<sub>WORD</sub>*. The participant read one sentence at a time, but words were presented separately one by one. The appearance of a new word triggered the previous word to disappear, such that at every point in time, only one word was displayed on the screen. Hence, the sentence was never displayed as a whole, and words had to be grasped during their individual presentation times. As in the other TRT versions, the presentation time of each word corresponded to the duration of the word in the audio recording of the sentence. Words were presented at the spatial position they would have had if the entire sentence had been displayed. Thus, the first word appeared in the leftmost position, and the last word appeared in the rightmost position. Again, participants were allowed to start their responses before the sentence was completed.

*TRT<sub>CENTER</sub>*. The design of this test resembles that of the *TRT<sub>WORD</sub>*. Only one word at a time was presented to the reader. However, in this test, every word appeared in the same spatial position in the center of the screen.

*TRT<sub>MEMORY</sub>*. In this test, 13 sentence pairs (rather than 13 single sentences) were presented to the reader. The individual sentences of each pair were subsequently presented in the same manner as in the *TRT<sub>500</sub>*. When the first sentence was finished, it disappeared. Then, after a short pause (1,000 ms), the second sentence was presented at the same percentage of masking as the first sentence of the pair. Participants were not allowed to start their response to any of the sentences before the presentation of the second sentence was finished. The participants were requested to repeat the two sentences in the correct order. Unknown to the participants, the adaptive masking procedure and calculation of the threshold were based on the correctness of the response to the second sentence only. Also, the order in which the sentences were repeated did not actually influence the rating of the response.

**Reading Span (RSpan) Test.** We created a Dutch version of the RSpan test to measure people's verbal WM capacity. For this RSpan test, we conformed to the approach

presented in Andersson et al. (2001) because it incorporated the dual-task component discussed in the introduction.

To assess the RSpan, we presented participants with 12 sets of three to six sentences ( $3 \times 3$ ,  $3 \times 4$ ,  $3 \times 5$ , and  $3 \times 6$  sentences) in increasing set-size order. Sentences were in past tense and had a subject-verb-object structure consisting of five words – for example, “The writer read the newspaper.” Half of the sentences were semantically correct; the other half were nonsense sentences – for example, “The coffee closed the window.” In all cases, the semantic inconsistency occurred between subject and verb. After every sentence, the participant had to indicate whether it was semantically correct or absurd by saying “right” or “wrong,” respectively. At the end of every set, participants were asked to recall either all subjects or all objects in the set. Participants did not know which of the two (subjects or objects) they would have to repeat until the set was finished. This ensured that people would try to remember the entire sentence instead of focusing only on a certain part of it. Also, we ensured processing of the sentences’ content by asking for the semantic judgments. Together, these two tasks formed the previously mentioned dual-task component and made this reading span test a “complex” WM test.

Sentences were presented in a three-step fashion, displaying one part of the sentence at a time (subject-verb-object). Articles were presented along with the subject and object nouns but did not need to be recalled later. The timing of the presentation was adopted from Andersson et al. (2001). Accordingly, sentence parts were displayed for 800 ms, with 75 ms between the parts and a 1,750-ms pause between the sentences. After the completion of a set, participants were given a maximum of 80 s to recall the requested sentence subjects or objects. In all cases, recall was finished in this interval. The outcome of the RSpan test was the total number of correctly recalled target words. The maximum score was 54.

**LDST.** We determined processing speed with an LDST (Van der Elst et al., 2006). This is a paper-and-pencil test, for which participants were given an answer key specifying nine unique letter-digit pairs, using all digits 1-9 in a random order. Beneath the key, rows of letters were printed, for which participants were requested to fill in the associated digits according to the key. The letters were printed in a nonsystematic order, and participants had to write down the digits following the rows. Left-handed participants were allowed to start at the right end of the rows if they wished. The test started with a set of 10 practice fill-ins. Then, the participants had 60 s to write down as many digits as possible. The outcome of the LDST was the absolute number of correctly substituted letters.

### **Procedure**

Participants were invited for a single test session that took about 2.5 hr, consisting of three testing blocks ( $2 \times 45$  min and  $1 \times 30$  min). Between the blocks, there were breaks of 15 min each. Within the testing blocks, all tests except the SRT and TRT tests were conducted at fixed positions. Pure-tone auditory thresholds, near vision, color blindness, and processing speed were assessed at the start of the test session. The RSpan test was administered directly after the first break. The rest of the test session comprised TRT and SRT tests. The order of the SRT and TRT test versions was counterbalanced across study participants following a Williams design (Williams, 1949). This type of Latin-square design balances the order of test administration across as few participants as possible while inhibiting carryover effects between the administered tests. The pair of Williams squares underlying this experiment were taken from Pezzullo (1999), where Williams designs are provided for those experiments that include two to 26 treatments, balanced for first-order carryover effects.

For each SRT masking condition (stationary vs. modulated), we used three lists with 13 sentences each and, thus, conducted three test runs. For the TRTs, we administered four

runs per test version. This procedure allowed us to exclude the first run from the data analyses in case of observed training effects (Zekveld et al., 2007). We used TRTs and SRTs averaged across test repetitions for the data analyses. TRT and SRT test stimuli were derived from the same corpus (Versfeld et al., 2000), but no sentence list was presented more than once. We balanced the allocation of sentence lists among tests and across study participants for the TRT and SRT tests separately.

After completing a TRT, SRT, or RSpan test, participants were asked to rate how taxing and how pleasant they found the test. We were interested in the subjective evaluation of the test difficulty because a clinically applicable test should be feasible and not too annoying. Also, the TRT should be comparably taxing as the SRT. We used seven-point scales ranging from 1 (*not taxing at all*) to 7 (*very taxing*) and from 1 (*very annoying*) to 7 (*very pleasant*).

### **Apparatus**

For study administration, we used a Dell Optiplex 755 PC (Intel Core 2 Duo CPU, E7300 at 2.66 GHz) with a Dell 1703 FP (17-in.) monitor (Dell Computer Corporation). The auditory stimuli were presented using the PC's internal soundcard (SoundMAX Integrated Digital HD Audio) and Sennheiser HDA 200 headphones (Sennheiser Electronic). Calibrations were done with a Brüel&Kjær Artificial Ear (Type 4152) and a Brüel&Kjær 2260 Observer to conform to the ISO 389 standard.

### **Data Analysis**

All analyses were performed using SPSS statistical package 15.0 for Windows. We checked the data of the RSpan, LDST, SRT, and TRT tests for univariate outliers and normality in terms of distribution by inspecting their histograms, mean and median values, SDs, and values for skewness and kurtosis.

Practice effects have been reported for SRT and TRT tests that are administered several times in a row. For instance, Zekveld et al. (2007) observed significantly poorer results for the first test run than for all subsequent runs of the TRT test. Hence, we examined practice effects in our data with repeated measures analyses of variance (ANOVAs), applying Bonferroni corrections for multiple comparisons.

After eliminating practice runs, we analyzed the test-retest reliability of the TRTs and SRTs by means of intraclass correlation coefficients (ICCs) and the standard error of measurement (SEM). We then calculated Pearson correlations between all TRT test versions, SRTs, RSpan, LDST, age, PTA, and level of education. We also computed partial correlations between the same tests, controlling for age. To gain further insights into the association of the TRTs and SRTs with the cognitive measures, we performed additional regression analyses. We created association models controlling for the following potential confounders: age, PTA, sex, and level of education. We also created prediction models for the  $SRT_{MOD}$  and the  $SRT_{STAT}$  to examine which factors most accurately explained the interindividual variances observed in these tests. This analysis served to identify the most relevant nonauditory contributors to speech comprehension in noise.

Finally, we looked into differences in the subjective scores between the tests. For this, we used ANOVAs, applying Bonferroni corrections for multiple comparisons.

## Results

The normality check of the test results revealed that the data of one participant deviated by at least 3 SDs from the mean on all TRT tests. Because these tests are central to the current study, we decided to exclude all data of this participant. Therefore, the analyses reported in this article are based on the data of 55 people ( $N = 55$ ). The examinations did not disclose any

abnormally distributed variables<sup>1</sup>. Accordingly, no data transformations needed to be applied. Histograms including normal curves for the TRT tests are given in Figure 2.

Investigation of the practice effects in TRT and SRT tests revealed that the SRT<sub>MOD</sub> was the only test that showed a clear practice effect with substantially better performance from the second test run onward (difference run1 – run2 and run1 – run3,  $p < .001$ ; difference run2 – run3,  $p = 1.0$ ). Therefore, we used the mean only of the second and third SRT<sub>MOD</sub> test in the analyses. None of the TRT versions showed a practice effect. We assume that the practice effects within the TRT tests were leveled out because the order in which the versions were conducted was counterbalanced across study participants.

We calculated the ICCs of the SRT and TRT tests, measuring absolute agreements with a two-way random model. Because we had several trials for every test (three per SRT and four per TRT), we provide average-measures ICCs in Table 2. An ICC of at least 0.7 generally is considered satisfactory (De Vet, Terwee, Knol, & Bouter, 2006; Scientific Advisory Committee of the Medical Outcomes Trust, 2002). The SEM is defined as  $SD * \sqrt{(1 - ICC)}$ , which indicates the variability of an individual's scores in repeated measurements, assuming that there are no practice effects.

We provide the outcomes of the reliability analysis together with test score means, ranges, and other descriptive data in Table 2. The results of the TRT<sub>ORIGINAL</sub> and the TRT<sub>500</sub> were similar to those found in other studies (George et al., 2007; Zekveld, Kramer, & Festen, 2011). The scores on each of the new TRT tests were higher (i.e., poorer) than for the TRT<sub>ORIGINAL</sub>, indicating that the participants needed to see a higher percentage of unmasked text to be able to reproduce 50% of the sentences correctly. Also the interindividual variance

---

<sup>1</sup> Due to software settings in the RSpan test, 37 subjects obtained slightly underestimated scores. Scores were underestimated by up to 5 points. Based on the data collected for the remaining 18 participants and data collected in independent studies (34 participants), we were able to estimate corrected scores for these first 37 participants. Corrected scores are reported in this paper. The analyses give the same pattern of results when we include the uncorrected data.

in test scores was larger for the new test versions (see also Figure 2). Furthermore, the new TRT tests had a better test–retest reliability than the original version. The latter was the only TRT variant with an ICC below the 0.7 criterion. Despite its relatively low ICC score, the TRT<sub>ORIGINAL</sub> has a low SEM score, caused by its rather low SD as compared with the SDs of the other TRT tests.

### Correlation Analyses

Unadjusted Pearson correlation coefficients among TRTs, SRTs, RSpan, LDST, age, PTA, and level of education are presented in Table 3. Speech understanding in modulated noise was strongly correlated with all TRT versions, both cognitive measures, and all other variables. For speech comprehension in stationary noise, fewer relationships were statistically significant. The SRT<sub>STAT</sub> correlated significantly with the TRT<sub>CENTER</sub>, the TRT<sub>WORD</sub>, the SRT<sub>MOD</sub>, and education (all  $p$ s < .01) but also with the original TRT ( $p$  < .05).

All TRT versions were negatively associated with RSpan, varying from  $r = -.36$  ( $p$  < .01, TRT<sub>ORIGINAL</sub>) to  $r = -.52$  ( $p$  < .01, TRT<sub>CENTER</sub>). This negative correlation actually represents a positive relationship between the variables: Better performance on the TRT test (reflected by a lower TRT score) was associated with better performance on the RSpan test (reflected by a higher RSpan score). The correlation coefficient between the TRT<sub>ORIGINAL</sub> and the RSpan was slightly lower than that between the RSpan and the new TRTs. Thus, the new TRT versions are more strongly associated with WM capacity than is the original test. A similar pattern held for the relationships between TRT tests and processing speed. The TRT<sub>ORIGINAL</sub> was not associated with the LDST, whereas three of the new tests were (all but the TRT<sub>WORD</sub>).

Additionally, the new TRT versions were more strongly correlated with age than was the original test. Higher age was associated with poorer performance (higher scores) on the

TRT tests. The highest correlation with age was found for the TRT<sub>500</sub>. Age is known to be a strong confounder in relationships involving speech understanding in adverse conditions (George, Festen, & Houtgast, 2006; Helfer & Freyman, 2008) or cognitive tests (Hällgren et al., 2001; Wingfield, Tun, & McCoy, 2005). Therefore, we examined the partial correlation coefficients between the variables controlling for age (see Table 4).

When controlling for age, the SRT<sub>MOD</sub> was no longer significantly correlated with RSpan or LDST. Equally, most correlations between the TRT versions and the RSpan disappeared, as did all correlations with the LDST. Three TRT versions were still significantly associated with the RSpan—namely, TRT<sub>CENTER</sub>, TRT<sub>WORD</sub>, and TRT<sub>MEMORY</sub>. It is important to note that the correlations between the TRTs and the SRTs remained, although they were slightly weaker. Thus, speech and text reception showed a solid relationship, even when controlling for age. This is consistent with the results observed by Zekveld et al. (2007).

Speech understanding also correlated with educational level, as did processing speed; however, none of the other factors did. A higher level of education was associated with better performance on the SRT tests and with higher processing speed. There were no significant correlations between PTA and the SRTs when controlling for age, which is the result of including only those individuals with normal hearing. Thus, the small differences in hearing sensitivity were completely governed by age.

## **Regression Analyses**

*Associations of text and speech reception with the cognitive variables.* A shortcoming of correlation analyses is that only one parameter at a time can be controlled for when examining the relationship between two variables. To get more detailed insights into the relationships between the different variables, and controlling for all confounders simultaneously, we performed linear regression analyses. We created association models to

examine the relationship of each TRT and SRT version with each of the cognitive measures separately and in greater detail. For this purpose, we used one TRT or SRT version at a time as the outcome measure (dependent variable) and either the RSpan or the LDST as the central determinant—that is, resulting in 10 TRT models (5 TRT versions with RSpan + 5 TRT versions with LDST) and 4 SRT models (2 SRT versions with RSpan + 2 SRT versions with LDST).

We examined the confounding effects of age, sex, level of education, and PTA (PTA being used for the SRT models only) on each of the 14 associations specified above. For all regression analyses, we coded education into a dichotomous variable of higher education (higher professional or university) and lower education. This was motivated by the fact that level of education did not show a linear relationship with the outcome measures.

We created the models by adding confounders one at a time according to the size of their confounding effect. This means that the strongest confounder was added first, and the weakest confounder was added last. When a new variable had been added to the model, the confounding effects of all remaining potential confounders were examined again. A variable needed to change the regression coefficient of the central determinant by at least 10% in order to be regarded as a confounder at all.

Essentially, the results of the 14 association models replicated the results of the correlation analysis, which is why we do not provide a table displaying the results. Overall, age was the predominant confounder in the relationships among the examined factors. In the TRT models, age was actually the only significant confounder. Consistent with the correlation analysis,  $TRT_{\text{CENTER}}$ ,  $TRT_{\text{WORD}}$ , and  $TRT_{\text{MEMORY}}$  remained significantly associated with the RSpan when controlling for age. All other observed associations between the TRTs and both cognitive measures were significant only when not controlling for age.

In the models that had an SRT as the outcome measure, level of education was the second strongest confounder after age. PTA was a confounder only for the relationship between SRT<sub>STAT</sub> and RSpan. However, also in the SRT models, the observed relationships were not significant anymore when controlling for age.

We also investigated whether age, sex, education, or PTA (PTA being used for the SRT models only) were effect modifiers in the associations between any of the TRT or SRT versions with the cognitive measures. In cases of observed effect modifications, we applied stratification to examine the associations within subgroups of our study population. Because of reduced group sizes in the stratified analyses, we considered a  $p$  value below .1 to be statistically significant in these analyses ( $p < .05$  in the other association analyses).

Effect modifications were found for age and education in the models including the TRT<sub>CENTER</sub>. There was a significant association of both RSpan and LDST, with the TRT<sub>CENTER</sub> only for people with a lower education. Age was an effect modifier only in the model with TRT<sub>CENTER</sub> and LDST. We created three subgroups of age using the variable's tertile cutoffs to further examine this effect. This way, we arrived at three groups: young ( $n = 18$ , range = 18–35 years), middle-aged ( $n = 19$ , range = 36–51 years), and old ( $n = 18$ , range = 52–78 years). The LDST was only significantly associated with TRT<sub>CENTER</sub> in the oldest group.

*Prediction of speech reception thresholds.* As a second step in the linear regression analysis, we created prediction models for the SRT<sub>MOD</sub> and the SRT<sub>STAT</sub> to examine which factors most accurately explained the interindividual variances observed in these tests in order to identify the most relevant nonauditory contributors to speech comprehension in noise. Variables included as potential predictors were the TRT versions, RSpan, LDST, education, age, sex, and PTA. The outcome measure was either the SRT<sub>MOD</sub> or the SRT<sub>STAT</sub>. Only one TRT version at a time was included in the group of potential predictors. Therefore, 10

prognostic models in total (i.e., 2 SRT versions  $\times$  5 TRT versions) needed to be created. We used a backward selection procedure for the model constructions, meaning that all potential predictors were initially included in the model. Then, based on a principle of least significance, variables were excluded one by one until only predictors were left that had a significance level of, at most, .1. These remaining predictors are presented for each of the 10 models in Table 5. For each model, the table states the total variance in SRT accounted for by that model's significant predictors together (adj.  $R^2$  total). The table also shows the variance in SRT that each predictor would account for on its own in a model including no other predictors (adj.  $R^2$  p. pred). We report adjusted  $R^2$  values as an indicator of the proportion of variance explained rather than the regular  $R^2$ , which tends to overestimate the explained variance.

Age was the strongest predictor of  $SRT_{MOD}$  in all models. Beyond age, the TRTs were the only measures with predictive power. An exception to this is the model including the  $TRT_{500}$ , in which level of education was also a significant, though weak, predictor. Of all models, the model including the  $TRT_{500}$  accounted for most of the variance in the  $SRT_{MOD}$  – namely, 50.5%. The model with the  $TRT_{ORIGINAL}$  accounted for 47.3% of the variance in  $SRT_{MOD}$ . All models explained at least 43% of the interindividual variance.

For the  $SRT_{STAT}$ , the strongest predictor was level of education, but overall, the amount of explained variance in  $SRT_{STAT}$  was low. Because education was coded into a dichotomous variable, its coefficient indicates the mean difference in  $SRT_{STAT}$  scores between people with higher education and people with lower education. Thus, people with higher education performed better on the  $SRT_{STAT}$ . Of all TRT versions, the  $TRT_{CENTER}$  best predicted the  $SRT_{STAT}$ . Together with education, it accounted for 28.7% of the variance observed in the  $SRT_{STAT}$  scores. Furthermore,  $TRT_{ORIGINAL}$ ,  $TRT_{WORD}$ , and  $TRT_{500}$  significantly contributed to explaining the interindividual variance in  $SRT_{STAT}$  – again,

together with education. Thus, TRT<sub>MEMORY</sub> was the only TRT that did not predict variance in the SRT<sub>STAT</sub>.

### **Analysis of the Subjective Measures**

The analysis of the subjective measures revealed that the TRT tests were equally pleasant as the SRT tests but were more pleasant than the RSpan test. An exception was the TRT<sub>MEMORY</sub>, which was comparably pleasant as the RSpan. Furthermore, the TRT tests (except the TRT<sub>MEMORY</sub>) were not experienced as more taxing than the SRT<sub>MOD</sub> and were only slightly more taxing than the SRT<sub>STAT</sub>. The RSpan test was considered more taxing than all other tests. These findings are of interest when considering clinical application of the TRT test, as it should not annoy or tire patients more than any alternative test.

### **Discussion**

The aim of this study was to make the TRT a real visual analogue to the SRT, especially to the more ecologically valid SRT<sub>MOD</sub> (Kramer et al., 1996). Observations from previous research (Kramer et al., 2009) motivated the assumption that this aim would likely be achieved by strengthening the associations of the TRT with WM capacity and processing speed. These abilities are considered relevant cognitive components of speech comprehension (Akeroyd, 2008). Therefore, a secondary aim of this research was to design new versions of the TRT test that would depend more strongly on WM capacity and processing speed.

In an unadjusted correlation analysis, we observed highly significant correlations of speech comprehension in modulated noise with all TRT versions and both cognitive variables. This supports the general modality-independent association between the SRT<sub>MOD</sub> and these measures. The SRT<sub>MOD</sub> also correlated with age, education, and PTA. In contrast, the SRT<sub>STAT</sub> was only associated with the TRT<sub>CENTER</sub>, TRT<sub>ORIGINAL</sub>, TRT<sub>WORD</sub>, and education.

The analysis confirmed that we succeeded in developing TRT versions that are more strongly associated with the cognitive measures. All new TRT versions showed higher correlations with the RSpan than did the TRT<sub>ORIGINAL</sub>. Three of the new versions were also correlated with the LDST, whereas the TRT<sub>ORIGINAL</sub> was not. Furthermore, correlations of the new TRT tests with age were higher than the correlation of the original test with age. Especially, the TRT<sub>500</sub> correlated relatively strongly with age.

When controlling for age in the correlation analysis, the LDST no longer correlated with any of the TRT versions, although the presentation times of the sentences were shortened in all new versions compared with the original test. Potentially, verbal tests that rely heavily on processing speed are more strongly associated with both TRT and SRT, such as lexical decision or rhyme judgment tests (Lunner, 2003). The RSpan remained significantly associated with three of the TRT tests – that is, the TRT<sub>CENTER</sub>, the TRT<sub>WORD</sub>, and the TRT<sub>MEMORY</sub> – when controlling for age. These were the TRT versions that clearly contained a memory component. They allowed the reader to read only one word at a time or required keeping two sentences in mind before responding. Remarkably, these WM capacity-related TRT versions showed the weakest associations with the SRT<sub>MOD</sub>. This fact indicates that WM capacity was not an essential determinant of speech comprehension in noise in the current study. Consistently, neither RSpan nor LDST were associated with speech comprehension in noise when controlling for age. On the contrary, the correlations between TRTs and SRTs remained significant. Given that our study included participants with normal hearing (according to PTAs) from a wide age range, these results indicate a solid relationship between masked text and speech reception, which is almost independent of age. This finding confirms the value of the TRT as a visual measure of the modality-overlapping cognitive components of speech comprehension in noise. The subjective data also showed that participants did not

experience any of the TRT tests except the TRT<sub>MEMORY</sub> as less pleasant or more taxing than the SRT tests. This finding is noteworthy regarding the clinical applicability of the test.

The vanishing of associations between the RSpan and the SRTs when controlling for age clearly suggests age-related declines in WM capacity and, thus, a stronger age dependency of the RSpan than that observed for the TRT. We also examined whether age was an effect modifier in the associations between the RSpan and the SRTs, using the following age groups: young (18–35 years), middle-aged (36–51 years), and old (52–78 years). No significant effect modifications were found, indicating that the association between RSpan and speech comprehension in noise does not differ for people in different age groups.

A predictive regression analysis revealed the contribution of the TRT tests and the other examined variables to the prediction of the SRTs. Age and all TRTs were significant predictors of the SRT<sub>MOD</sub>. Also, level of education was a weak predictor of the SRT<sub>MOD</sub>, but only in the model including the TRT<sub>500</sub>. The SRT<sub>STAT</sub> was predicted by education and four of the five TRT versions (TRT<sub>ORIGINAL</sub>, TRT<sub>500</sub>, TRT<sub>CENTER</sub>, and TRT<sub>WORD</sub>). None of the other variables significantly predicted speech understanding in noise.

Previous studies found significant correlations or predictive value of the RSpan test for speech comprehension in noise, including both stationary and modulated noise (Foo et al., 2007; Hällgren et al., 2001; Lunner, 2003; Rudner et al., 2008). In the cases where age was controlled for, these studies also revealed weaker associations between RSpan and speech comprehension in noise. What distinguishes these studies from the current one is that people with mild to moderate hearing loss (PTA around 45 dB), with and without hearing aids, were included (Foo et al., 2007; Hällgren et al., 2001; Lunner, 2003; Rudner et al., 2008). Also, the mean age of the study participants in those studies was between 65 and 70 years. Thus, the participants were substantially older than those in the current study. Furthermore, the age ranges in these studies were much more limited than in our population, which restricted the

observable age-related declines in WM capacity in the previous studies and, thus, the confounding effect of age.

In terms of the ELU model (Rönnberg et al., 2008), explicit processing resources such as WM capacity are tapped only if the signal cannot be interpreted by implicit phonological processing, a so-called mismatch situation. One could argue that the SRTs did not evoke mismatch effects in our sample of individuals with normal hearing. However, we assessed SRTs at an intelligibility level of 50%. This means that the participants could reproduce only half of the sentences correctly. Accordingly, mismatches occurred in at least 50% of the sentences. The presence of mismatch effects was also confirmed by the associations of RSpan and  $SRT_{MOD}$  in our analyses when not controlling for age. Also, researchers have previously reported that the relationship between cognitive variables such as RSpan and speech perception is more pronounced for amplitude-modulated noise maskers than for stationary maskers (Rönnberg et al., 2010).

The reading span test that we used was newly created for this study; this Dutch version has not been used in research before. The test was carefully matched to the test described in Andersson et al. (2001) in terms of sentence complexity, manner and timing of stimulus presentation, and task. As stated earlier, the most important feature of a valid WM capacity test is that it includes a dual task, which requires the tested person to not only store but also process incoming information. This condition was fulfilled in the RSpan test used in the current study. The results of the RSpan test were normally distributed with a large variance in result scores, which indicates that the test was sensitive to interindividual differences in WM capacity. Thus, this Dutch version of the test can be considered a valid tool for measuring WM capacity.

An evident conclusion from the current study is that the TRT test and the RSpan test measure essentially different nonauditory components of speech understanding in noise. We

found a nearly age-independent association of the TRT with the SRT, indicating that it is a valuable measure for the nonauditory part of speech understanding in adverse conditions. We assume that the TRT test measures acquired linguistic processing skills that are relevant in speech comprehension to a larger extent than the cognitive ability assessed by the RSpan test. Regarding the design of the TRT and the SRT test, these linguistic skills should be related to a person's ability to fill in informational gaps in a verbal stimulus, independent of the input modality. The reading span test might represent a more general WM capacity for the processing of verbal input, which apparently is more sensitive to age-related declines. However, Zekveld, Rudner, et al. (2011) observed that both the TRT and RSpan tests are associated with the ability to ignore irrelevant semantic context during speech perception, indicating that both tests are related to more specific, context-dependent speech perception ability. Therefore, we suggest that researchers use both tests in parallel to examine the broad range of an individual's cognitive processing capabilities.

The data from this study did not clearly identify one of the TRT test versions as the best visual analogue of the SRT test. Both the TRT<sub>ORIGINAL</sub> and the TRT<sub>500</sub> were good predictors of the SRT<sub>MOD</sub>. However, the TRT<sub>500</sub> accounted for slightly more of the variance observed in the SRT<sub>MOD</sub>, had a higher test–retest reliability than the TRT<sub>ORIGINAL</sub>, and was more strongly associated with age. In this sense, the TRT<sub>500</sub> implies an improvement, as compared with the TRT<sub>ORIGINAL</sub>. The gain in explained SRT variance by the new versions compared with the original test was small, which could indicate that the amount of variance in the SRT that can be accounted for by a visual test such as the TRT is limited.

The TRT that best predicted the SRT<sub>STAT</sub> was the TRT<sub>CENTER</sub>. The TRT<sub>CENTER</sub> appears to be different from the other TRTs in several ways. Its association with level of education was higher than for the other TRTs. This might account for its relation to the SRT<sub>STAT</sub>, which was also highly associated with education. It remains unclear why education influenced the

performance of these two tests in particular. Furthermore, the TRT<sub>CENTER</sub> was the only TRT for which we observed effect modifications in its associations with the cognitive measures. The TRT<sub>CENTER</sub> scores of the older participants and participants with a lower degree of education depended on WM capacity and processing speed to a greater extent than those who had a higher education or were younger. The correlation of the TRT<sub>CENTER</sub> with the RSpan was also higher than for any other TRT test. The question arises regarding whether the manner of stimulus presentation could have influenced this outcome. Both in the TRT<sub>CENTER</sub> and the RSpan test, the words of the sentences are presented in the center of the screen, which could alleviate the temporal processing demands in these tests. In the other TRT tests, sentence words appear at different spatial positions.

Future research studies including different versions of the TRT test will provide further insights into the characteristics of the different tests and their relationship with speech understanding in noise in other study populations and also other kinds of speech comprehension tasks.

### Conclusions

We have successfully developed TRT tests that showed stronger associations with speech comprehension in noise, WM capacity, processing speed, and age than an earlier version of this test. In contrast to previous research, speech understanding in noise was not associated with WM capacity or processing speed when controlling for age. Accordingly, the TRT test versions that were most strongly related to the SRT<sub>MOD</sub> – namely, the TRT<sub>500</sub> and the TRT<sub>ORIGINAL</sub> – were not the ones with evident connections to the examined cognitive measures. This finding stresses the fundamental difference of the abilities relevant for speech comprehension in noise that are measured by the TRT and RSpan tests. The RSpan test is a dual-task measure reflecting an individual's general WM capacity available for both storage

and simultaneous processing of verbal input. Alternatively, the TRT test more specifically assesses the ability to make sense of fragmentary verbal information. The latter appears to be independent of the input modality, as was revealed by a robust association between masked text recognition and speech-in-noise perception over and above the influences of confounding factors such as age and education. This result confirms the value of the TRT test as an instrument for assessing the nonauditory modality-overlapping linguistic abilities relevant for speech comprehension in adverse listening conditions. Thus, the TRT can be considered an eligible visual analogue of the SRT test. Moreover, the tests are equivalent regarding the subjective evaluation of how taxing and pleasant it is to perform the tests.

We identified the  $TRT_{500}$  as the best predictor of the  $SRT_{MOD}$  in addition to age and level of education, together accounting for 51% of the observed variance in the  $SRT_{MOD}$ . The  $TRT_{ORIGINAL}$  also significantly predicted the  $SRT_{MOD}$ , but compared with the  $TRT_{500}$ , it had a lower test–retest reliability. A weakness of the  $TRT_{500}$  is its relatively weak association with speech comprehension in stationary noise, but we consider the outcomes for the  $SRT_{MOD}$  more relevant for clinical purposes due to its higher ecological validity. The  $SRT_{STAT}$  was best predicted by level of education and the  $TRT_{CENTER}$ , but the amount of variance in  $SRT_{STAT}$  accounted for was generally low.

### **Acknowledgments**

Funding for this research was provided by Foundation Het Heinsius-Houbolt Fonds. We would also like to acknowledge the support received from the Swedish Research Council, the technical assistance and software development provided by Hans van Beek, and the efforts of the study participants.

### References

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*(Suppl. 2), S53–S71.
- Andersson, U., Lyxell, B., Rönnberg, J., & Spens, K.-E. (2001). Cognitive correlates of visual speech understanding in hearing-impaired individuals. *Journal of Deaf Studies and Deaf Education, 6*, 103–116.
- Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology, 50*, 371–384.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*, 417–423.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.
- Baddeley, A., Logie, R., Nimmosmith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24*, 119–131.
- Bailey, I. L., & Lovie, J. E. (1980). The design and use of a new near-vision chart. *American Journal of Optometry and Physiological Optics, 57*, 378–387.
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 60*, 223–233.
- British Society of Audiology. (2004). *Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels*. Retrieved from [www.thebsa.org.uk/docs/RecPro/PTA.pdf](http://www.thebsa.org.uk/docs/RecPro/PTA.pdf).

- Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*, 499–509.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology-Learning Memory and Cognition*, *9*, 561–584.
- De Vet, H. C.W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, *59*, 1033–1039.
- Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*, 538–546.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*, 1725–1736.
- Foo, C., Rudner, M., Rönnerberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, *18*, 618–631.
- Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42*(Suppl. 1), S77–S85.
- George, E. L. J., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*, 2295–2311.

- George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T. (2007). Auditory and non-auditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, *121*, 2362–2375.
- Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2001). Cognitive effects in dichotic speech testing in elderly persons. *Ear and Hearing*, *22*, 120–129.
- Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, *29*, 87–98.
- Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology*, *47*, 287–295.
- Humes, L. E. (1991). Understanding the speech-understanding problems of the hearing impaired. *Journal of the American Academy of Audiology*, *2*, 59–69.
- Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., & Strausner, L. E. (2007). Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effect of presentation rate. *Journal of Speech, Language, and Hearing Research*, *50*, 283–303.
- Ishihara, S. (1989). *The series of plates designed as a test for colour-deficiency*. Tokyo, Japan: Kanehana Trading.
- Janse, E. (2009). Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, *125*, 2361–2373.
- Jolles, J., Houx, P. J., Van Boxtel, M. P. J., & Ponds, R. W. H. M. (1995). *Maastricht Aging Study: Determinants of cognitive aging*. Retrieved from [www.np.unimaas.nl/maas/moreinfo/MAAS\\_PB\\_intro.pdf](http://www.np.unimaas.nl/maas/moreinfo/MAAS_PB_intro.pdf).

- Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Tobi, H. (1996). The relationships between self-reported hearing disability and measures of auditory disability. *Audiology, 35*, 277–287.
- Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of the Text Reception Threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology, 50*, 507–515.
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology, 42*(Suppl. 1), S49–S58.
- Lunner, T., & Sundewall-Thoren, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology, 18*, 604–617.
- Nachtegaal, J., Smit, J. H., Smits, C., Bezemer, P. D., van Beek, J. H., Festen, J. M., & Kramer, S. E. (2009). The association between hearing status and psychosocial health before the age of 70 years: Results from an Internet-based national survey on hearing. *Ear and Hearing, 30*, 302–312.
- Pezzullo, J. C. (1999). *Latin squares for constructing "Williams Designs," balanced for first-order carry-over (residual) effects*. Retrieved from <http://statpages.org/latinsq.html>.
- Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*(Suppl. 2), S72–S82.
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*, 593–608.

- Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification*, 10, 29–59.
- Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech, Language, and Hearing Research*, 29, 146–154.
- Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43–52.
- Rönnberg, J. (1990). Cognitive and communicative function: The effect of chronological age. *European Journal of Cognitive Psychology*, 2, 253–275.
- Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, 42, 68–76.
- Rönnberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speech-reading and cognitive function. *Journal of Speech and Hearing Research*, 32, 725–735.
- Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, 47(Suppl. 2), S99–S105.
- Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, 12, 263–269.
- Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal of Psychology*, 50, 405–418.

- Rudner, M., Foo, C., Sundewall-Thoren, E., Lunner, T., & Rönnberg, J. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *International Journal of Audiology, 47*(Suppl. 2), S91–S98.
- Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56*, 139–152.
- Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life measurements: Attributes and review criteria. *Quality of Life Research, 11*, 193–205.
- Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence, 37*, 283.
- Stenfelt, S., & Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology, 50*, 385–393.
- Stephens, D., & Kramer, S. E. (2010). *Living with hearing difficulties: The process of enablement* (1st ed.). West Sussex, United Kingdom: Wiley-Blackwell.
- Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Letter Digit Substitution Test: Normative data for 1,858 healthy participants aged 24–81 from the Maastricht Aging Study (MAAS): Influence of age, education, and sex. *Journal of Clinical and Experimental Neuropsychology, 28*, 998–1009.
- Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of speech reception threshold. *The Journal of the Acoustical Society of America, 107*, 1671–1684.

- Wechsler, D. (1955). *Manual for the Wechsler Intelligence Scale*. New York, NY: The Psychological Corporation.
- Wechsler, D. (1981). *Manual for the Wechsler Intelligence Scale—Revised*. New York, NY: The Psychological Corporation.
- Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, Series A: Physical Sciences*, 2, 149–168.
- Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, 7, 175–182.
- Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology*, 18, 548–558.
- Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*, 14, 144–148.
- World Health Organization. (2011). *Grades of hearing impairment*. Retrieved from [www.who.int/pbd/deafness/hearing\\_impairment\\_grades/en/](http://www.who.int/pbd/deafness/hearing_impairment_grades/en/).
- Zekveld, A. A., George, E. L., Kramer, S. E., Goverts, S. T., & Houtgast, T. (2007). The development of the Text Reception Threshold test: A visual analogue of the Speech Reception Threshold test. *Journal of Speech, Language, and Hearing Research*, 50, 576–584.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, 32, 498–510.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., van Beek, J. H. M., & Rönnerberg, J. (2011). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and Hearing*, *32*, e16–e35.

**Table 1.** Demographic variables and hearing acuity of the study participants.

Age (years)	N	% men	Edu	PTA <sub>TEST</sub>	PTA <sub>NON</sub>	
< 30	11	36.4	5.9	3.5	4.4	
30-39	11	18.2	6.2	3.3	5.1	
40-49	13	23.1	6.3	6.2	8.9	
50-59	12	33.3	5.9	10.4	12.3	
>= 60	8	25.0	4.5	11.7	15.9	
<b>All</b>	44.0	55	27.3	5.9	6.8	9.0

*Note.* Edu = the average level of education on an ordinal scale from 1 (*primary*) to 7 (*university*); PTA = the average hearing loss in dB HL at octave frequencies 0.5 – 4 kHz for the test ear (PTA<sub>TEST</sub>) and the nontest ear (PTA<sub>NON</sub>).

**Table 2.** Test scores for the speech reception thresholds (SRTs), text reception thresholds (TRTs), and cognitive tasks.

<b>Variable</b>	<b><i>M</i></b>	<b><i>Mdn</i></b>	<b><i>SD</i></b>	<b>Range</b>	<b>Min</b>	<b>Max</b>	<b>ICC</b>	<b><i>SEM</i></b>
SRT <sub>MOD</sub> (dB SNR)	-7.6	-7.8	1.6	8.4	-10.6	-2.2	.4	1.2
SRT <sub>STAT</sub> (dB SNR)	-3.1	-3.1	.8	3.6	-4.5	-.9	.5	.6
TRT <sub>ORIGINAL</sub> (%)	56.0	55.9	2.8	13.8	50.2	64.0	.6	1.7
TRT <sub>500</sub> (%)	58.9	58.9	4.4	18.3	50.8	69.1	.8	1.9
TRT <sub>CENTER</sub> (%)	59.7	58.9	5.4	26.1	48.7	74.8	.8	2.5
TRT <sub>WORD</sub> (%)	62.7	61.0	5.4	22.5	53.3	76.0	.8	2.2
TRT <sub>MEMORY</sub> (%)	64.3	63.4	6.3	27.1	54.4	81.5	.8	2.9
RSpan	19.7	19.7	6.1	30.0	4.0	34.0		
LDST	36.6	36.0	6.6	29.0	21.0	50.0		

*Note.* Intraclass correlations (ICCs), and standard errors of measurement (*SEMs*) are not applicable for the cognitive tests because they were administered only once. For the SRTs and TRTs, higher scores indicate poorer results. For the cognitive measures, higher scores imply a better result. Min = minimum score; Max = maximum score; SNR = signal-to-noise ratio; RSpan = reading span; LDST = Letter–Digit Substitution Test.

**Table 3.** Pearson correlations (two-tailed).

Variable	TRT ORIGINAL	TRT 500	TRT CENTER	TRT WORD	TRT MEMORY	PTA	age	edu	LDST	RSpan	SRT <sub>STAT</sub>
SRT <sub>MOD</sub>	.48**	.59**	.46**	.46**	.48**	.40**	.64**	-.40**	-.40**	-.42**	.47**
SRT <sub>STAT</sub>	.33*	.26	.49**	.36**	.23	.20	.24	-.41**	-.17	-.21	
RSpan	-.36**	-.39**	-.52**	-.41**	-.49**	-.40**	-.52**	.30*	.50**		
LDST	-.22	-.41**	-.38**	-.19	-.28*	-.35**	-.55**	.42**			
education	-.16	-.18	-.34*	-.19	-.10	-.17	-.29*				
age	.31*	.50**	.41**	.36**	.36*	.62**					
PTA	.06	.17	.22	.14	.05						

*Note.* Negative correlations with SRTs and TRTs indicate a positive relationship of the variables due to the fact that better performance on the SRT and TRT tests was reflected by lower scores.

\* Sign. at the .05 level. \*\* Sign. at the .01 level.

**Table 4.** Partial Pearson correlations, controlling for age (two-tailed).

Variable	TRT ORIGINAL	TRT 500	TRT CENTER	TRT WORD	TRT MEMORY	PTA	edu	LDST	RSpan	SRT <sub>STAT</sub>
SRT <sub>MOD</sub>	.38**	.41**	.28*	.32*	.35*	.00	-.30*	-.08	-.14	.43**
SRT <sub>STAT</sub>	.28*	.17	.44**	.30*	.16	.07	-.37**	-.05	-.10	
RSpan	-.25	-.17	-.39**	-.28*	-.39**	-.11	.18	.30*		
LDST	-.07	-.19	-.20	-.01	-.11	-.02	.32*			
education	-.08	-.05	-.26	-.10	.01	.01				
PTA	-.18	-.21	-.06	-.12	-.25					

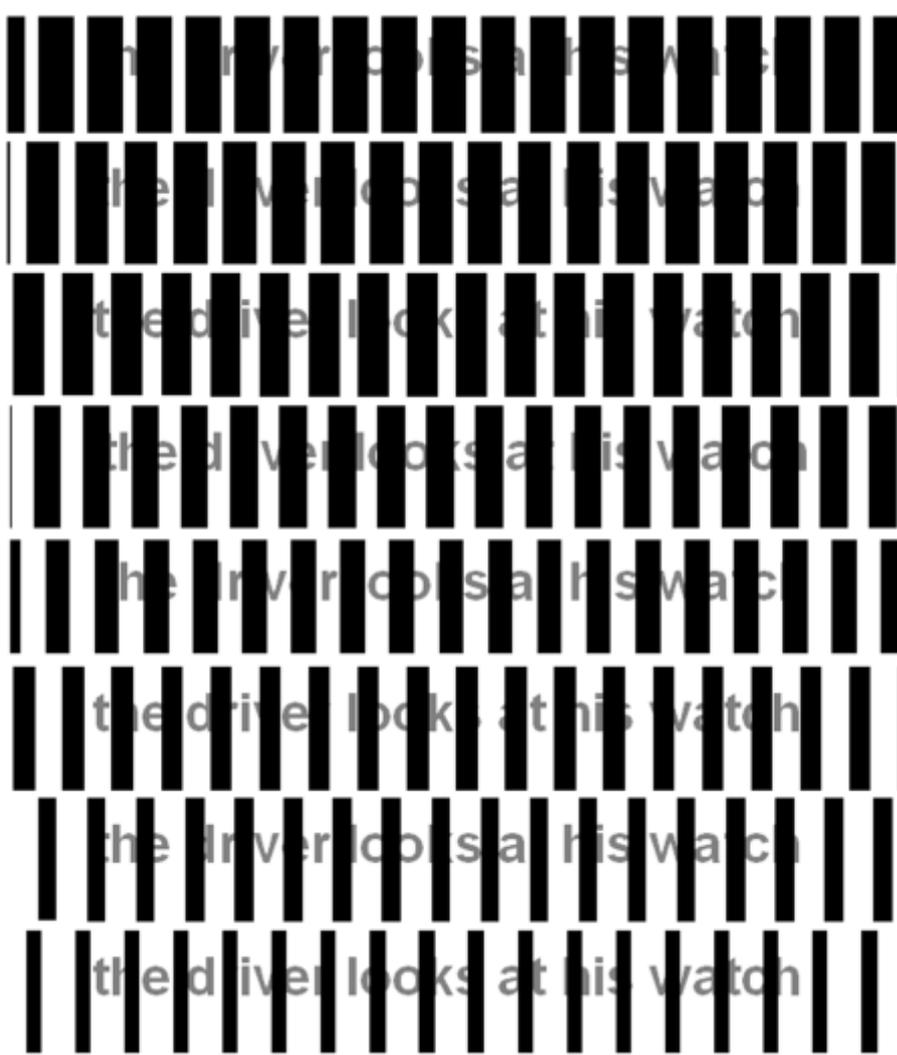
*Note.* Negative correlations with SRTs and TRTs indicate a positive relationship of the variables due to the fact that better performance on the SRT and TRT tests was reflected by lower scores.

\* Sign. at the .05 level. \*\* Sign. at the .01 level.

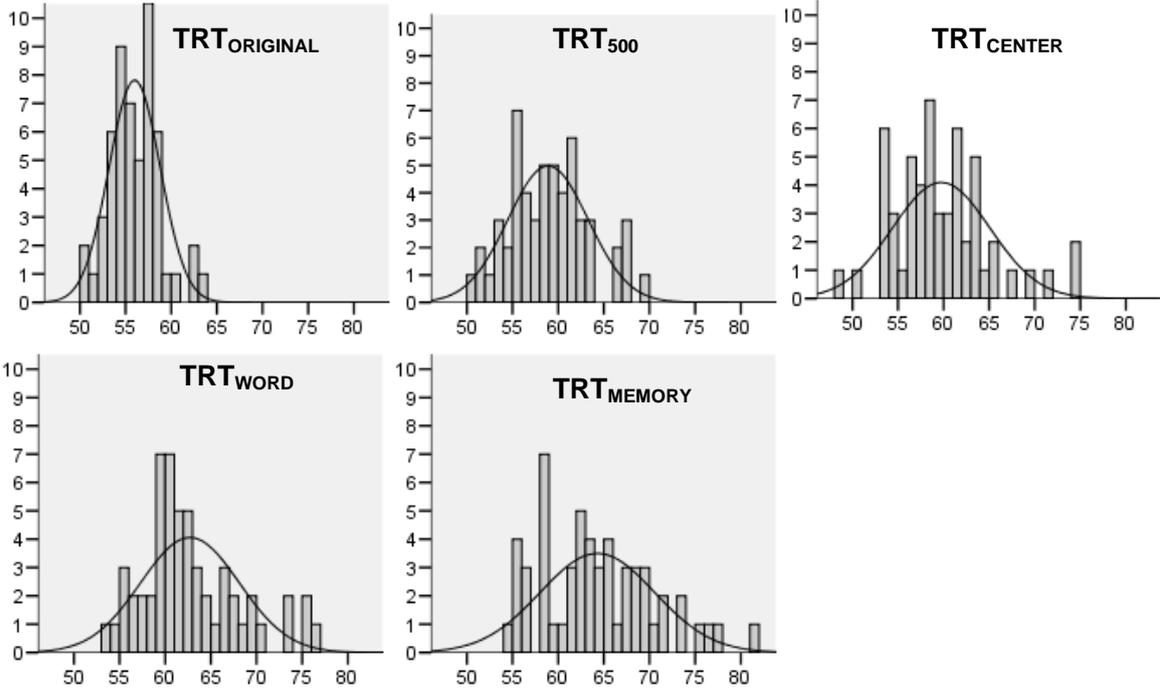
**Table 5.** Prediction of SRTs by TRTs, cognition, sex, PTA, age, and level of education (low vs. high).

SRT <sub>MOD</sub>							SRT <sub>STAT</sub>						
Predictor	<i>B</i>	$\beta$	95% CI	<i>p</i>	Adj. $R^2$ p. pred	Adj. $R^2$ total	Predictor	<i>B</i>	$\beta$	95% CI	<i>p</i>	Adj. $R^2$ p. pred	Adj. $R^2$ total
<b>model with TRT<sub>ORIGINAL</sub></b>													
Age	.062	.539	[.038 – .086]	.000	.394	.473	Edu	-.606	-.355	[-1.024 – -.189]	.005	.128	.207
TRT <sub>ORIGINAL</sub>	.182	.312	[.060 – .304]	.004	.216		TRT <sub>ORIGINAL</sub>	.090	.305	[.018 – .162]	.015	.095	
<b>model with TRT<sub>500</sub></b>													
Age	.049	.423	[.026 – .078]	.000	.394	.505	Edu	-.628	-.367	[-1.084 – -.213]	.004	.128	.173
TRT <sub>500</sub>	.139	.375	[.053 – .221]	.001	.340		TRT <sub>500</sub>	.046	.244	[-.001 – .093]	.054	.051	
Edu	-.553	.164	[-1.211 – .104]	.097	.045								
<b>model with TRT<sub>CENTER</sub></b>													
Age	.062	.537	[.036 – .088]	.000	.394	.432	TRT <sub>CENTER</sub>	.065	.423	[.029 – .102]	.001	.222	.287
TRT <sub>CENTER</sub>	.074	.241	[.004 – .143]	.038	.199		Edu	-.487	-.285	[-.892 – -.083]	.019	.128	
<b>model with TRT<sub>WORD</sub></b>													
Age	.062	.540	[.037 – .088]	.000	.394	.445	Edu	-.592	-.346	[-1.007 – -.177]	.006	.128	.219
TRT <sub>WORD</sub>	.080	.265	[.014 – .146]	.018	.197		TRT <sub>WORD</sub>	.050	.323	[.012 – .087]	.010	.112	
<b>model with TRT<sub>MEMORY</sub></b>													
Age	.062	.533	[.037 – .086]	.000	.394	.457	Edu	-.649	-.379	[-1.084 – -.213]	.004	.128	.128
TRT <sub>MEMORY</sub>	.075	.288	[.019 – .131]	.010	.215		(TRT <sub>MEM</sub> )					(.036)	

*Note.* *B* = nonstandardized regression coefficient;  $\beta$  = standardized regression coefficient; CI = confidence interval; adj.  $R^2$  p. pred = adjusted  $R^2$ -values per predictor; adj.  $R^2$  total = adjusted  $R^2$  values for the whole model.



**Figure 1.** Text reception threshold (TRT) test sentence, “The driver looks at his watch,” at 8 different levels of unmasked text from 28% unmasking (top row) to 70% unmasking (bottom row), with a masking decrement of 6% per row. The figure is reproduced from Zekveld et al. (2007).



**Figure 2.** Histograms with normal distribution curves for each of the TRT test versions.