Development of the Word Auditory Recognition and Recall Measure: A Working Memory Test for Use in Rehabilitative Audiology

Sherri L. Smith,[1,2] M. Kathleen Pichora-Fuller,[3,4,5,6], and Genevieve Alexander[1]

Objectives: The purpose of this study was to develop the Word Auditory Recognition and Recall Measure (WARRM) and to conduct the inaugural evaluation of the performance of younger adults with normal hearing, older adults with normal to near-normal hearing, and older adults with pure-tone hearing loss on the WARRM.

Design: The WARRM is a new test designed for concurrently assessing word recognition and auditory working memory performance in adults who may have pure-tone hearing loss. The test consists of 100 monosyllabic words based on widely used speech-recognition test materials. The 100 words are presented in recall set sizes of 2, 3, 4, 5, and 6 items, with 5 trials in each set size. The WARRM yields a word-recognition score and a recall score. The WARRM was administered to all participants in three listener groups under two processing conditions in a mixed model (between-subjects, repeated measures) design. The between-subjects factor was group, with 48 younger listeners with normal audiometric thresholds (younger listeners with normal hearing [YNH]), 48 older listeners with normal thresholds through 3000 Hz (older listeners with normal hearing [ONH]), and 48 older listeners with sensorineural hearing loss (older listeners with hearing loss [OHL]). The within-subjects factor was WARRM processing condition (no additional task or with an alphabet judgment task). The associations between results on the WARRM test and results on a battery of other auditory and memory measures were examined.

Results: Word-recognition performance on the WARRM was not affected by processing condition or set size and was near ceiling for the YNH and ONH listeners (99 and 98%, respectively) with both groups performing significantly better than the OHL listeners (83%). The recall results were significantly better for the YNH, ONH, and OHL groups with no processing (93, 84, and 75%, respectively) than with the alphabet processing (86, 77, and 70%). In both processing conditions, recall was best for YNH, followed by ONH, and worst for OHL listeners. WARRM recall scores were significantly correlated with other memory measures. In addition, WARRM recall scores were correlated with results on the Words-In-Noise (WIN) test for the OHL listeners in the no processing condition and for ONH listeners in the alphabet processing condition. Differences in the WIN and recall scores of these groups are consistent with the interpretation that the OHL listeners found listening to be sufficiently demanding to affect recall even in the no processing condition, whereas the ONH group listeners did not find it so demanding until the additional alphabet processing task was added.

Conclusions: These findings demonstrate the feasibility of incorporating an auditory memory test into a word-recognition test to obtain measures of both word recognition and working memory simultaneously. The correlation of WARRM recall with scores from other memory measures is evidence of construct validity. The observation of correlations between the WIN thresholds with each of the older groups and recall scores in certain processing conditions suggests that recall depends on listeners' word-recognition abilities in noise in combination with the processing demands of the task. The recall score provides additional information beyond the pure-tone audiogram and word-recognition scores that may help rehabilitative audiologists assess the listening abilities of patients with hearing loss.

Key words: Aging, Audiologic rehabilitation, Hearing loss, Speech understanding, Working memory.

(Ear & Hearing 2016;37;e360–e376)

[1]Audiologic Rehabilitation Laboratory, Auditory Vestibular Research Enhancement Award Program, Veterans Affairs Medical Center, Mountain Home, Tennessee, USA; [2]Department of Audiology and Speech Language Pathology, East Tennessee State University, Johnson City, Tennessee, USA; [3]Department of Psychology, University of Toronto, Mississauga, Ontario, Canada; [4]Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada; [5]Rotman Research Institute, Baycrest Hospital, Toronto, Ontario, Canada; and [6]Linneaus Centre HEAD, Linköping University, Sweden.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

INTRODUCTION

The comprehension of spoken language involves cognitive abilities, including working memory, processing speed, attention, executive function, and compensatory use of semantic context and linguistic knowledge (e.g., Wingfield & Stine-Morrow 2000; Schneider et al. 2002; Rönnberg 2003; Pichora-Fuller & Singh 2006; Rönnberg et al. 2008; Rönnberg et al. 2013; Souza & Arehart 2015). Although various cognitive abilities are involved in speech understanding, much of the research in this area has focused on working memory (for reviews see Akeroyd 2008; Pichora-Fuller 2013).

Working memory refers to the system involved in both the temporary storage of information and the use of that information to perform cognitive tasks (Baddeley & Hitch 1974; Daneman & Carpenter 1980; Shah & Miyake 1999). Working memory is capacity limited, meaning that there is a finite amount of working memory resources available for an individual to process and store information related to a current cognitive task, such as speech comprehension (e.g., Just & Carpenter 1992). Working memory is important for speech understanding in everyday conversations because listeners must interpret the meaning of running speech as it unfolds by relating it to knowledge stored in long-term memory as well as to the information that was heard previously and that will be heard next (Daneman & Carpenter 1980, 1983; Pichora-Fuller et al. 1995). The amount of working memory capacity available for speech understanding varies across individuals (interindividual differences). In addition, the amount of working memory capacity allocated to processing speech, as opposed to storing speech in memory, varies under different listening conditions within an individual (intraindividual differences; Pichora-Fuller 2007).

Interindividual Differences in Working Memory

Individual differences in working memory capacity may be important in explaining functional differences between listeners.

0196/0202/2016/376-e360/0 • Ear & Hearing • Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

<zdoi; 10.1097/AUD.0000000000000329>

For example, comparing two individuals with good pure-tone thresholds who are comprehending speech heard in quiet, the individual who has a larger working memory capacity will use a smaller proportion of that capacity to process (i.e., encode) the speech signal compared with an individual who has a lower working memory capacity. After the incoming signal has been processed by two such listeners, the listener with higher working memory capacity will have relatively more "spare" working memory capacity available to allocate to remembering and comprehending what has been heard (see Pichora-Fuller 2007 for review; also see Rudner et al. 2011b; Rönnberg et al. 2014). Older listeners, even those with normal pure-tone thresholds, may need to allocate more working memory resources to understanding speech than would be needed by younger adults with normal hearing because age-related declines in suprathreshold auditory processing abilities may make it more difficult for them to encode the incoming information (Pichora-Fuller et al. 1995). For individuals with pure-tone hearing loss or suprathreshold auditory processing deficits, listening may consume even more working memory resources for processing speech than would be consumed by counterparts with better audiograms or normal auditory processing, even in relatively favorable listening conditions. Consequently, even assuming that there are no interindividual differences in total working memory capacity, individuals with auditory deficits who need more working memory resources to process speech may have poorer memories for what they heard compared with those with better hearing.

Studies examining interindividual differences in working memory report correlations that show a trend for those with higher working memory to have better speech-recognition performance in more adverse listening conditions (unaided and aided) relative to those with lower working memory (e.g., Akeroyd 2008; Besser et al. 2013; Souza & Arehart 2015). Importantly, when amplification has been used to overcome the effects of signal inaudibility, then cognitive factors have explained as much and sometimes even more of the variance in speech recognition in noise by older adults than could be explained by pure-tone thresholds alone (e.g., Rönnberg et al. 2008, 2010; Besser et al. 2013; Humes et al. 2013). Thus, the cognitive abilities of older adults seem relevant to audiologic rehabilitation, especially because listening to and comprehending speech in many challenging everyday situations can be a complex task that requires more cognitive processing than would be required in relatively simple speech recognition tasks (Kiessling et al. 2003; Pichora-Fuller 2006a; Baldwin & Ash 2011). In the context of rehabilitation, if audiologists were to measure an aspect of the cognitive processing of speech, with an auditory working memory test for speech for example, then that information may be a useful supplement to the information provided by typical speech-recognition scores.

Intraindividual Differences in Working Memory

In addition to interindividual differences in working memory capacity, it may be of interest to know about intraindividual differences in how working memory resources are allocated across various listening conditions and tasks within an individual. Intraindividual differences in working memory for speech can be thought of as the resource allocation between storage and encoding/processing within one individual's given capacity under different listening conditions. In other words, there is a trade-off between the resources allocated for encoding/processing and storage that varies with the demands of the listening condition for a given individual (Pichora-Fuller 2007).

Notably, hearing loss, background noise, or any algorithm that distorts the speech waveform may exacerbate the processing demands required during listening (where listening refers to hearing with intention and attention [Kiessling et al. 2003]). Because working memory capacity is limited, as processing demands increase and more working memory resources are allocated to listening, there will be a reduction in the resources that remain available for comprehension or other cognitive processing of the information that has been heard, and there also will be a reduction in how much of the information that has been heard becomes stored in memory (Rabbitt 1991; Gordon-Salant & Fitzgibbons 1997; Pichora-Fuller 2003; Pichora-Fuller & Singh 2006; Wingfield et al. 2006; Wingfield & Tun 2007). Indeed, research suggests that there can be detrimental effects on recall (i.e., storage) when the quality of the signal is reduced in some manner, including by reducing the presentation level of auditorily presented speech materials (e.g., Baldwin & Ash 2011; Amichetti et al. 2013), by adding noise (e.g., Rabbitt 1968), or by altering the rate of speech (e.g., Piquado et al. 2012).

Pichora-Fuller (2006a) argued that clinicians might find working memory span measures to be useful for assessing interindividual differences to identify individuals who may be candidates for certain rehabilitative interventions, for example, working memory may be relevant to the choice of slow versus fast hearing aid compression (Gatehouse et al. 2006a, b; Lunner & Sundewall-Thorén 2007; Souza & Sirow 2014). She also argued that measuring intraindividual differences in working memory capacity may be clinically useful to determine if interventions have had a positive effect on working memory. For example, if amplification purportedly makes the encoding and processing auditory signals easier relative to the unaided condition for a given individual, then there should be more working memory resources available for storage for the individual wearing amplification, with an increase in working memory span scores being a positive hearing-aid outcome (e.g., Doherty & Desjardins 2015). Moreover, intraindividual differences in working memory span scores may be useful in evaluating the outcomes of rehabilitative audiology because span measures may be more sensitive to change than traditional word-recognition measures, especially when word recognition is at or near ceiling (e.g., Pichora-Fuller 2003; Pichora-Fuller 2006a, b; Pichora-Fuller & Singh 2006). A prerequisite to measuring condition-specific intraindividual differences and/or outcomes of audiologic rehabilitation is to develop a clinically feasible test of auditory working memory.

Working Memory Measures

Various working memory measures have been used in studies that have examined the role of working memory in speech understanding, including but not limited to the following: (1) the auditory working memory subtest from the Woodcock Johnson cognitive test battery (Woodcock et al. 2001), (2) a Simon-Says measure (Humes & Floyd 2005), (3) digit span forward (DSF) and backward from the Wechsler Adult Intelligence Scale (Wechsler 1981, 1987, 1997), (4) reading span (e.g., Daneman & Carpenter 1980; Rönnberg et al. 1989), and (5) listening

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

span (Pichora-Fuller et al. 1995), which is an auditory counterpart of the reading span measure that uses Revised Speech in Noise sentences as target stimuli (Bilger 1984). These measures vary in the stimuli used (i.e., speech versus nonspeech), the presentation modality (i.e., visual versus auditory), the processing task employed, and test administration time. Of all the working memory measures, perhaps the most commonly used working memory measure in audiology research has been the reading span measure (e.g., Lunner 2003; Foo et al. 2007; Rudner et al. 2007; Arehart et al. 2013; Besser et al. 2013; Ng et al. 2013). In fact, a recent review by Besser et al. summarized 21 studies that examined the role of verbal working memory in speech recognition in noise, 16 studies of which used reading span.

One of the most common and oldest versions of the reading span test involves the visual presentation of 100 sentences in set sizes of 2, 3, 4, 5, and 6 sentences with 5 trials in each set size (Daneman & Carpenter 1980). The sentences are presented visually one at a time and the task of the subject is to process each sentence by reading it aloud and/or making a judgment about it such as whether it is semantically possible or not (e.g., "The ocean shone brightly over the clear, blue sun."). At the end of each trial, a recall prompt appears and the individual then recalls verbally the final word in each of the sentences in the trial. The premise of the test is that as the amount of working memory allocated to processing the semantic appropriateness of the sentence increases there will be a corresponding decrease in the number of words that can be successfully stored and recalled. Across studies using various working memory measures, the reading span measure has been the most successful in correlating with language comprehension measures (Daneman & Merikle 1996) and also with various speech recognition measures (Akeroyd 2008; Besser et al. 2013). The reading span measure also has been the measure of choice when the focus has been on interindividual differences in working memory in relation to speech understanding (for reviews see Akeroyd 2008 and Besser et al. 2013), but it would not be able to provide information about intraindividual differences in working memory for speech under different listening conditions given that it is presented in the visual modality.

Modality of Working Memory Testing

The selection of visual or auditory materials for tests of working memory seems to have been guided by two different theoretical views. According to one view, working memory is characterized by a single-component, general-purpose working memory capacity system that is common to and shared by all cognitive tasks (e.g., Broadbent 1958; Daneman & Carpenter 1980; Engle et al. 1992, 1999). This theoretical view supports the use of visually presented stimuli (e.g., text). In many auditory studies, working memory tests were presented in the visual modality, presumably to avoid "contamination" by the effects of hearing loss (e.g., Souza 2012; Besser et al. 2013).

An alternative theoretical view is that working memory is a multicomponent system that is differentially accessed depending upon the cognitive task being accomplished, or in other words, that there are domain-specific working memory systems (Baddeley & Logie 1999; Cocchini et al. 2002; Baldwin & Ash 2011). A variation on this view is that the allocation of working memory resources may be modality dependent, as well as being task dependent. Recent debates have questioned the desirability of testing working memory in the visual modality to isolate modality-general abilities rather than testing working memory directly in the auditory modality when auditory functioning is of clinical interest (see Dillon et al. 2014; McFarland & Cacace 2014; Moore & Ferguson 2014 for discussion). Accordingly, it would be more meaningful to use speech stimuli presented in the auditory modality to measure working memory when examining the role of working memory in spoken speech understanding. In some studies, various working memory tests have been presented in the auditory modality, presumably because the authors assumed that there could be modality-specific factors and/or they were interested in examining the perceptual processing/storage trade-off for speech (i.e., intraindividual differences) in ecologically relevant conditions that matched those in which speech understanding performance was being measured (Besser et al. 2013). By understanding the interplay of auditory and cognitive processing within an individual across a range of listening and task conditions, the audiologist would gain useful insights into how these factors combine to influence everyday communication functioning (e.g., Baldwin & Ash 2011; Besser et al. 2013).

In summary, the audiology research to date has focused on interindividual differences in working memory capacity, and the reading span measure has been the most popular measure used to quantify working memory capacity. Although the reading span measure provides information about interindividual differences, the test may not be appropriate for assessing how the allocation of working memory resources depends on hearing abilities and/or test conditions of functional or ecological relevance that may influence intraindividual variations in performance. The reading span test presented in the visual modality is not as well suited to measuring the outcomes of hearing-related interventions as an auditory working memory span test might be insofar as an auditory test should be more sensitive to improvements in auditory processing that could reduce the allocation of working memory resources during listening and leave more spare capacity available for further processing such as comprehension and memory of what was heard. Although auditory working memory measures exist (e.g., Woodcock-Johnson, listening span by Pichora-Fuller et al. 1995, etc.), they are not well suitable for use in audiology clinical settings to test listeners with hearing loss because they are limited in terms of the type of stimuli used or because of the lengthy time needed for test administration.

Measuring Working Memory in Audiology

Measures of working memory have been developed and standardized for use in the general population (e.g., digit span, Woodcock-Johnson), and some tests have been developed for use in laboratory settings (e.g., Pichora-Fuller et al. 1995), or in other languages for use in listeners with hearing loss (e.g., Auditory Inference Span test by Rönnberg et al. 2011; Sentence-final Word Identification and Recall test by Ng et al. 2013); however, the existing measures have not been developed specifically to assess auditory working memory for speech in an audiology clinical setting. To fill this gap, we developed an auditory working memory test, the Word Auditory Recognition and Recall Measure (WARRM). The WARRM employs the basic working memory span test paradigm combined with a word-recognition test procedure. The test stimuli are well-known materials that control for parameters that may

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

influence word recognition (e.g., talker differences, lexical variables, etc.). The materials are recorded on compact disc (CD) to allow easy administration at calibrated presentation levels chosen to ensure the audibility of the test materials. The WARRM can yield a word-recognition score and a working memory recall score. Thus, in one test, the two scores will inform the rehabilitative audiologist about general word-recognition abilities in addition how much cognitive processing was involved regarding those words as reflected by the recall score. Furthermore, the item-by-item recall score takes into account the word that was recognized. Here, we describe the development of the WARRM test and provide a report of the initial WARRM test results obtained from younger and older adults with normal to near-normal audiograms and from older adults with clinically significant sensorineural hearing loss. This inaugural study sets the stage for further studies leading to a standardized test suitable for use by rehabilitative audiologists in clinical settings.

MATERIALS AND METHODS

Participants

Three groups of listeners participated in the present study. The groups included 48 younger listeners with normal hearing (YNH) defined as pure-tone thresholds ≤20 dB HL for 250 to 8000 Hz (M = 22.8 years, SD = 2.7 years, range 18 to 38 years; 17 males, 31 females), 48 older listeners with normal hearing (ONH) defined as pure-tone thresholds ≤25 dB HL at 250 to 3000 Hz (M = 66.9 years, SD = 5.1 years, range = 60 to 84 years; 8 males, 40 females), and 48 older listeners with sensorineural hearing loss (older listeners with hearing loss [OHL]; M = 70.3 years, SD = 5.8 years, range = 60 to 85 years; 46 males, 2 females). An analysis of variance (ANOVA) revealed a significant difference in age (3.4 years) between the two older listener groups, $F(1, 95) = 9.1$, $p < .005$. The highest level of education reported (in years) on average by the participants was 16.1 (SD = 2.1), 15.6 (SD = 2.8), and 14.4 (SD = 1.9) for the YNH, ONH, and OHL listeners, respectively. An ANOVA revealed a significant group difference, $F(2, 143) = 6.8$, $p = .001$, such that the YNH and ONH had similar levels of education, but both groups had more years of education than the OHL group. The average audiograms for the three listener groups are illustrated in Figure 1. For each group at each frequency, the mean between-ear threshold differences were all ≤2.3 dB with an ANOVA confirming no significant ear by frequency interaction. For these reasons, only the right ear thresholds are plotted in Figure 1.

To be included in the study, participants had to have better than 50% correct word-recognition performance in quiet (Northwestern University Auditory Test No. 6 [NU-6; Tillman & Carhart 1966; Department of Veterans Affairs 2006]) so that floor effects with test stimuli would be avoided. All participants had to pass the Mini-Mental State Exam (MMSE) cognitive screening tool based on criterion provided by the established age and education norms (Folstein et al. 1975; Crum et al. 1993). None of the participants had a comorbid health condition (e.g., neurological condition, substance abuse, retrocochlear pathology, etc.) that potentially could interfere with study participation. Identification of comorbid health conditions was based on a chart review (for OHL listeners) and/or case history interview (for YNH and ONH listeners).

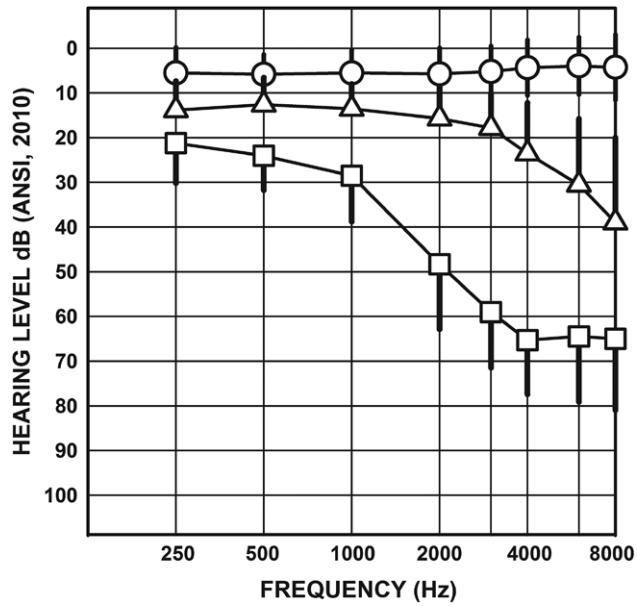

Fig. 1. The mean right-ear audiogram of the younger normal hearing listeners (circles), older listeners with normal hearing (triangles), and older listeners with hearing loss (squares) are illustrated. The error bars represent 1 SD.

Convenience sampling was used to recruit participants prospectively from the local community via flyers, from our audiology clinic, and from our participant recruitment registry. The YNH listeners were recruited from East Tennessee State University and the area community; none were audiology students. The ONH and OHL listeners were recruited from the local area and from the Mountain Home, Tennessee Veterans Affairs (VA) Medical Center Audiology clinic. Some participants were recruited from our participant recruitment registry that contains a list of individuals who are interested in being contacted for research. All participants had an equal chance of participating if they met the inclusion criteria.

Materials

In this study, audio recordings of speech stimuli were used in the following behavioral measures: WARRM, a second version of the Words-In-Noise test (Wilson 2003; Wilson et al. 2003), auditory digit span, and auditory free recall. The materials are from a 582-word corpus spoken by VA female speaker #2 (Wilson et al. 2008) that included the NU-6, the CID W-22s (Hirsh et al. 1952), the PB-50s (Egan 1948) word lists, and the monosyllabic digits (see also the Supplemental Digital Content, http://links.lww.com/EANDH/A303). Using this corpus ensured that there would not be test differences attributable to a talker effect (e.g., Kreul et al. 1969). In addition, as described below, we also administered a test of visual free recall, a vocabulary test, and the Montreal Cognitive Assessment (MoCA; Nasreddine et al. 2005).

Word Auditory Recognition and Recall Measure • In many ways, the WARRM test paradigm is an auditory counterpart of the reading span measure (Daneman & Carpenter 1980). The WARRM contains 100 monosyllabic words organized in set sizes of 2, 3, 4, 5, and 6 words with 5 trials in each set size. The WARRM stimuli, which are presented in quiet, consist of a carrier phrase ("You will cite") followed by a target word. The selection of these 100 target words is detailed in the

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

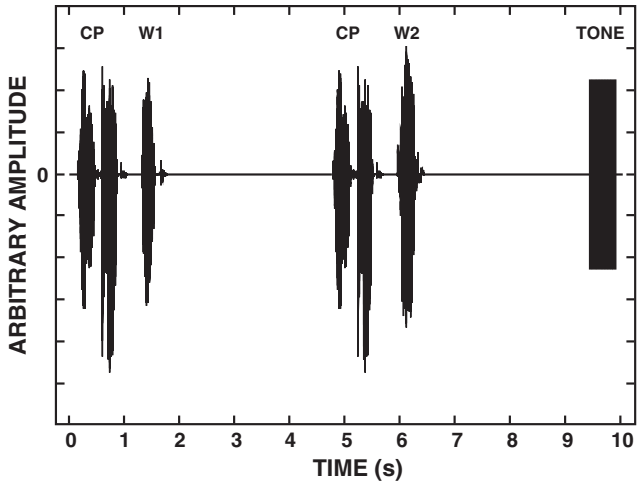

Fig. 2. A waveform representing a trial for set size 2. Tone = 500-msec, 500-Hz prompt tone. CP, carrier phrase; W1, word 1; W2, word 2.

Supplemental Digital Content (http://links.lww.com/EANDH/A303). The WARRM paradigm is illustrated in Figure 2 and is described as follows: the carrier phrase is followed by a 250-msec silent interval and then the target word with a 3000-msec silent interval between the target word offset and the onset of the next carrier phrase, and after the last target word in each trial, there is a 3000-msec silent interval and then a 500-msec, 500-Hz tone that serves as a recall prompt. This low-frequency alerting tone was used because of its relative immunity to pure-tone hearing loss. During the 3000-msec interstimulus intervals (ISIs), the participant repeated the target word (word-recognition task) and made a judgment about the word to engage cognitive processing. The words in the recall set were spoken by the participant after the recall prompt.

The processing task used in the WARRM ("alphabet" processing task) required the participant to report if the first letter of each target word belonged to the first half of the alphabet (A–M) or to the second half of the alphabet (N–Z). Accordingly, 50 of the target words started with letters from the first half of the alphabet and 50 target words started with letters from the second half of the alphabet. In the even numbered set sizes (2, 4, and 6), each trial contained an equal number of words beginning with letters from each half of the alphabet. In the odd numbered set sizes (3 and 5), there is an extra word from each half of the alphabet in each trial, but overall there is an equal distribution (also see Table 1 in Supplemental Digital Content (http://links.lww.com/EANDH/A277) for additional details about distribution of the words across trials). See Figures 1 and 2 in Supplemental Digital Content (http://links.lww.com/EANDH/A277) for details about the alphabetic frequency of the words in the test in reference to statistics for the written English language. In the present study, an alternative administration of the WARRM that did not include an additional processing task (referred to as "none" or the "no processing condition") was tested as a comparison condition.

The WARRM produces two scores. First, a word-recognition score is calculated in overall percent correct (across the 100 items) or as a function of set size. Second, a recall score is calculated, either in percent correct (overall or as a function of set size) or as a span score, which is a common way of scoring working memory tests. Following Daneman and Carpenter (1980), the span score corresponds to maximum set size for which at least three of the five trials in the set size are recalled correctly. If, for example, an individual correctly recalls 3 of 5 trials with set size 4, and less than 3 of 5 trials at higher set sizes, then the span score would be 4. Partial credit, however, is given for up to two correct trials in the next highest set size. For example, if an individual correctly recalls 1 trial in the next highest set size (e.g., set size 5 in this example), then partial credit of 0.33 (1 of 3) is added to the span score, for a total 4.33. If two trials are correctly recalled in the next highest set size (set size 5 in this example), then partial credit of 0.67 is given (or 2 of 3) for a span score of 4.67. Thus, the base number indicates the set size in which the individual "passes" (3 of 5 trials correct), and the number after the decimal indicates the proportion out of three trials in the next highest set size recalled correctly. Also noteworthy about recall scoring is that participants are given credit for correctly recalling the misperceived words during the recognition task (following Pichora-Fuller et al. 1995). If there is no response for a given word during recall testing, then the word is scored as incorrect. Finally, an additional score, referred to as the judgment score (in percent correct), is calculated only for the alphabet processing condition, but this score is expected to be nearly perfect and serves primarily to ensure that the listeners engage in the processing task. Like the recall score, participants are given credit for correctly judging any misperceived words during the recognition task.

Words-in-Noise Test with the VA Female Speaker #2 (WIN#2) • The original Words-In-Noise test (Wilson 2003; Wilson et al. 2003) consists of two lists of 35 NU-6 words presented in a 6-talker babble. The WIN paradigm involves the presentation of 5 unique words at each of 7 signal to noise ratios (S/N) from 24 to 0 dB in 4-dB decrements. The level of the babble is held constant and the level of the speech is varied. With WIN#2, the original WIN speech signals (carrier phrase *Say the word* and target words) were replaced by the corresponding stimuli recorded by the VA female speaker #2 using a different carrier phrase (*You will cite*). The same recording of the carrier phrase was concatenated with each of the 70 target words and mixed with the 70 babble segments from the original WIN. The listener's task is to repeat the target word and to guess when unsure. The 50% point (in dB S/N) is calculated using the Spearman-Kärber equation (Finney 1952). Unpublished data from our laboratory demonstrated that the 50% points of the WIN and the WIN#2 were similar, differing by <1 dB for YNH. None of the words used in the WARRM were the same as the words used in the WIN#2 test.

Digit Span • A modified audio version of the digit span subtest of the Weschler Adult Intelligence Scale (4th edition, WAIS-IV, Wechsler 2008) was recorded. Typically, the digit span task is administered in a face-to-face interview with the examiner presenting the digits (0 to 10) using live voice with about a 1-second ISI. The recorded version used the 9 monosyllabic digits (1 to 10, excluding the bisyllabic 0 and 7 that were used in the WAIS-IV) spoken by VA female speaker #2 that were compiled into a progressive sequence of digit sets with a 1-second ISI that terminated with a 500-msec, 500-Hz tone that signaled the listener to respond. The use of recorded materials provided a standardized protocol that controlled differences due to talker variables (including auditory and visual speech characteristics), presentation level, and timing differences that are not controlled in the standard face-to-face administration

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

of the test. Other than the recorded presentation, the standard procedures for administering and scoring the digit span were followed. The digit span test consists of three subtests including DSF, digit span backward (DSB), and digit span sequencing (DSS). The three digit span subtests differed in terms of the response that was required. In the DSF task, the listener is asked to repeat the digits in the order in which they were presented. In the DSB task, the listener is asked to repeat the digits in the reverse order in which they were presented. In the DSS task, the listener is asked to repeat the digits in ascending numerical order. In each subtest, a series of digit trials are administered, ranging from a set size with two digits in a trial up to set sizes with eight (DSB) or nine (DSF, DSS) digits in a trial. There are two trials for each set size, which makes up an item. One point is given for each trial correctly repeated. The test is terminated when an item score (total of two trials of equal set size) is zero. The total raw score possible for each subtest is 16.

Auditory Free Recall • A list of 15 monosyllabic words (without carrier phrases) were recorded (ISI = 2 seconds), terminating with a 500-msec, 500-Hz tone that prompted the participants to write down in a 3-min interval as many words as they could remember. The test was scored by summing the number of words that the listener correctly recalled. See Table 2 in Supplemental Digital Content (http://links.lww.com/EANDH/A277) for a list of the 15 words used in this test (adapted from Rabbitt 1968, 1991; Park et al. 1996).

Visual Free Recall • The VFR task is a visual counterpart to the AFR task and was included to provide a nonauditory comparison measure. For the VFR, a list of 15 words (Table 2 in Supplemental Digital Content, http://links.lww.com/EANDH/A277), which were different from the AFR words, was presented as text in a power point slide show (Microsoft) on a 38-cm computer screen (Dell 1908FP), one at a time, with a 1-second ISI. The words (72-point regular black Calibri lowercase font) were centered on a white background. Following the 15 words, a yellow slide with the word "RECALL" (72-point regular black Calibri uppercase font) was shown as the recall prompt. VFR recall and scoring procedures were identical to those in the AFR test (adapted from Rabbitt 1968, 1991).

Vocabulary • The participants were administered the 30-word vocabulary subtest of the WAIS-IV test according to the test administration and scoring procedures. In short, the examiner read the word to the participant, while the participant looked at the printed word in the stimulus booklet (starting with item 5, "apple"). The words were ordered from easy to difficult. If the participants cannot define the word *apple* to demonstrate good understanding, then they were asked to define earlier words on the list (all of our participants could define the word *apple*). The participants were asked to define each word, and their definitions were recorded verbatim in writing by the examiner on the score sheet. Each definition was scored with a 2 (i.e., good understanding of the word), a 1 (i.e., correct definition but lacking content), or a 0 (i.e., no understanding of the word even after query). The test was terminated after 3 consecutive scores of 0. The scores were summed and the total possible raw score was 57.

Montreal Cognitive Assessment • The MoCA screens for mild cognitive impairment by assessing attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. The maximum score is 30 points, with a score of ≥26 suggestive of normal cognitive function. The MMSE was used to determine eligibility for the study, but the MoCA score was included as a descriptor variable because it is more sensitive than the MMSE in detecting mild cognitive impairment (scores of 22 to 25; Nasreddine et al. 2005).

Procedures

The study was approved by the East Tennessee State University/VA Institutional Review Board and the VA Research and Development Committee before its initiation. All participants provided informed consent before commencing with study procedures. Testing was accomplished in 2, 90-min sessions with breaks included in each session. The participants were remunerated $20 per hour at the end of each test session. In session 1, following the informed consent process, the participants in the YNH and ONH groups were asked about their medical history to determine if a comorbid condition existed; a medical record chart review was performed to determine if comorbid conditions existed for the OHL participants. All participants then completed the MMSE. A standard audiologic evaluation was conducted, including an otoscopic examination, tympanometry and ipsilateral acoustic reflex threshold testing (Grason-Stadler, Model Tympstar), and pure-tone audiometry for octave frequencies 250 to 8000 Hz and interoctave frequencies of 3000 and 6000 Hz (Grason-Stadler, Model 61) using insert earphones (Etymōtic, Model ER-3A).

For the experimental portion of session 1, the participants completed the three subtests of the modified digit span test, WARRM (no judgment condition) in quiet, the vocabulary test, and either the auditory or visual free recall test. The test order of the experimental measures in session 1 was randomly assigned to participants. In session 2, the participants completed the WARRM (alphabet processing condition) in quiet, two 35-word lists of the WIN#2 (counterbalanced), either the auditory or visual free recall test (which ever was not presented in session 1), and the MoCA. The OHL participants had hearing aids, which they wore during the MMSE, MoCA, vocabulary, and visual free recall testing. All audio-recorded test materials were reproduced on a CD player (Sony, Model CDP-CE500), routed through the audiometer, and presented binaurally through insert earphones, while the participant was seated in the double-walled sound booth (IAC Acoustics). The materials were presented binaurally to represent an ecologically relevant condition of listening. The presentation levels for the recorded materials were 70 dB HL for listeners who had pure-tone averages (PTAs; re: 500, 1000, and 2000 Hz) ≤40 dB HL and 80 dB HL for listeners with PTAs ≥40 dB HL (American National Standards Institute 2010) and were chosen in attempt to ensure audibility of the auditory test materials.

RESULTS

The WARRM was the main experimental measure. The WARRM results (recognition and recall) are presented for the two processing conditions (no processing and alphabet processing). The judgement score was obtained only for the alphabet processing condition. Finally, the evaluation of the relations among the WARRM recall results, measures of memory, and the WIN#2 using correlational and/or hierarchical regression analyses are presented.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 1. The mean recognition performance in percent correct (and 1 SD) as a function of processing condition and set size are listed for each listener group

		YNH		ONH		OHL	
Processing	Set size	M	SD	M	SD	M	(SD)
None	2	100.0	0.0	98.1	4.5	70.4	22.0
	3	99.3	3.1	98.6	3.1	87.2	12.6
	4	99.6	1.7	98.5	2.9	82.5	14.7
	5	99.8	0.8	99.7	3.0	80.8	12.6
	6	97.2	2.6	96.4	3.5	79.9	12.1
	Overall	98.9	1.0	97.9	2.1	80.8	12.2
Alphabet	2	100.0	0.0	98.8	3.3	75.0	20.4
	3	99.9	1.0	99.3	2.1	90.7	9.2
	4	99.1	2.2	98.3	3.2	84.1	11.7
	5	99.9	0.6	98.6	3.4	84.2	13.8
	6	97.5	2.6	97.5	2.6	80.5	12.1
	Overall	99.0	1.2	98.5	2.6	83.1	11.3

OHL, older listeners with hearing loss; ONH, older listeners with normal hearing; YNH, younger listeners with normal hearing.

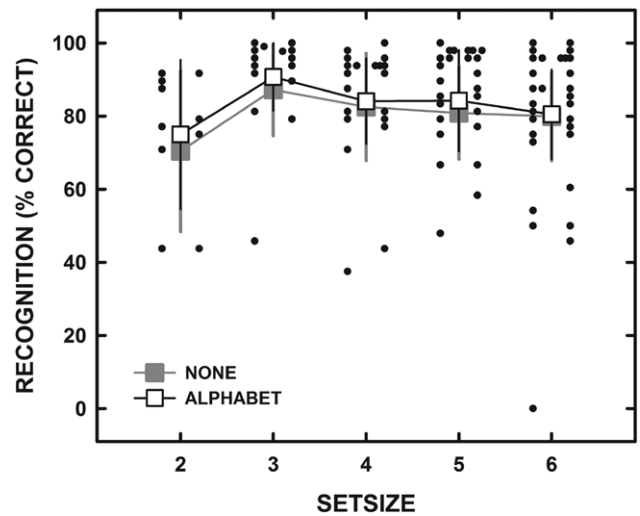

Fig. 3. The mean recognition performance (percent correct) is plotted as a function of set size for the older listeners with hearing loss. The large squares show the mean performance (with error bars representing 1 SD) for all the words in the set size for the no processing task (grey) and for the alphabet processing task (white). The small filled circles represent the mean performance for the individual words in each set size in the alphabet processing condition and are plotted on either side of the large symbols to avoid superimposition and improve graphical clarity.

Word-Recognition Performance

Table 1 provides the mean percent correct word recognition obtained in the two processing conditions, at each set size by each of the three listener groups. Guessing was encouraged, but if the participant made no response to a word during the recognition portion of the test, then the word was scored as incorrect. The YNH and ONH listeners had recognition scores that were at ceiling with little variability across set sizes in both processing tasks. In contrast, word-recognition performance by the OHL listeners was lower and varied more among the five set sizes. Figure 3 displays the mean recognition scores (large squares) for the OHL listeners as a function of set size and processing condition; the small circles show the recognition scores for the each word in each set size in the alphabet processing condition. Importantly, the data in the figure show that word-recognition accuracy is not negatively affected by adding the alphabet processing task and does not decline as memory load is increased with increasing set size. Also note that there are few words for which noticeably lower accuracy scores are achieved.

A repeated measures ANOVA was conducted*, with processing condition (none and alphabet) and set sizes (2, 3, 4, 5, and 6) as the within-subjects factors for the OHL listeners only. There was a main effect (with Greenhouse-Geisser corrections) of processing condition, $F(1, 47.0) = 12.6$, $p < .001$, $\eta_p^2 = 0.21$, and of set size, $F(4, 100.6) = 29.0$, $p < .001$, $\eta_p^2 = 0.38$, but no significant two-way interaction between processing task and set size. Post hoc analyses with Bonferroni corrections for multiple comparisons were conducted. For the main effect of processing condition (collapsed across set sizes), word-recognition accuracy was significantly better in the alphabet processing condition than in the no processing condition (by 2.3% or less than 3 words on this 100-item test). For the main effect of set size, compared with all other set sizes, word-recognition accuracy was worst for set size 2 and best for set size 3, with no differences amongst set sizes 4, 5, and 6. The interested reader is referred to Table 3 in Supplemental Digital Content (http://links.lww.com/EANDH/A277) for the mean performance for each word in each set size for the three listener groups.

* Another analysis was conducted when the percent correct data were arcsine transformed, but the raw scores are presented because the transformation did not change the statistical interpretation.

Judgment Score for the Alphabet Processing Condition

For the alphabet processing condition only, the participants were instructed to judge the first letter of the target word as belonging to the first half (A–M) or the second half (N–Z) of the alphabet. A judgment score was calculated as the percentage of the words that were judged correctly. As expected, the overall mean judgment scores for the test in the alphabet processing condition were at near-ceiling performance: 98.8% (SD = 0.7%) for the YNH, 98.7% (SD = 0.7%) for the ONH, and 96.9% (SD = 1.1%) for the OHL listeners. These scores indicate that the participants were engaged in processing the target stimuli because they were able to correctly indicate which half of the alphabet contained the first letter of the target word.

Recall

Recall in the two processing conditions for each listener group are presented for (1) overall percent correct recall across the 100 words, (2) percent correct recall by set size, and (3) working memory span.

Overall Percent Recall • The mean recall scores for each group and processing condition are listed in Table 2. For all groups, recall was poorer in the alphabet processing condition than in the no processing condition. Recall was best for the YNH group, followed by the ONH group, and was poorest for OHL group. A mixed model, repeated measures ANOVA was conducted to examine overall recall differences, with processing condition (none versus alphabet) as a within-subject factor and group (YNH, ONH, and OHL) as a between-subjects factor. The results (with Greenhouse-Geisser corrections) showed a main effect of processing condition, $F(1.0, 141.0) = 129.0$,

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 2. The mean recall performance in percent correct (and 1 SD) as a function of processing condition and set size are listed for each listener group

		YNH		ONH		OHL	
Processing	Set size	M	SD	M	SD	M	SD
None	2	99.6	2.0	98.1	4.9	94.8	8.5
	3	97.5	4.9	93.5	9.5	89.3	11.4
	4	97.9	3.8	90.7	9.5	80.7	14.8
	5	94.8	6.5	83.4	12.8	72.8	17.3
	6	84.9	11.9	71.3	15.1	60.1	15.9
	Overall	93.3	5.3	84.0	9.8	75.1	12.1
Alphabet	2	98.8	3.3	94.4	7.1	93.8	7.3
	3	95.8	5.6	89.6	12.2	86.4	11.1
	4	89.8	9.3	83.9	13.1	74.6	16.0
	5	84.9	11.0	72.8	14.7	65.3	15.8
	6	75.0	12.0	61.9	13.2	53.8	12.1
	Overall	86.0	7.6	76.5	10.5	69.7	11.2

OHL, older listeners with hearing loss; ONH, older listeners with normal hearing; YNH, younger listeners with normal hearing.

$p < .001$, $\eta_p^2 = 0.48$, and a main effect of group, $F(2,141) = 44.2$, $p < .001$, $\eta_p^2 = 0.39$, but no significant interactions. Post hoc analyses with Bonferroni corrections for multiple comparisons were significant ($p < .001$). Specifically, for the main effect of group collapsed across processing condition, (1) the YNH group had 9.4 and 17.3% better recall compared with the ONH and OHL listeners, respectively, and (2) the ONH group had 7.9% better recall than the OHL group. Collapsed across groups, recall in the no processing condition was 6.7% better than in the alphabet processing condition and the differences between processing conditions were similar for the YNH (7.3%), ONH (7.5%), and OHL (5.3%) groups.

Percent Recall Accuracy by Set Size • As anticipated, and as can be seen in Table 2, recall accuracy decreased as set size increased for all listener groups. A mixed model, repeated measures ANOVA was conducted with group (YNH, ONH, and OHL) as a between-subjects factor and with percent correct recall for each set sizes (2, 3, 4, 5, and 6) and processing condition (none and alphabet) as within-subjects factors. The results (with Greenhouse-Geisser corrections) showed a main effect on recall of processing condition—$F(1,141.0) = 136.1$, $p < .001$, $\eta_p^2 = 0.49$—set size—$F(2.81,396.3) = 450.6$, $p < .001$, $\eta_p^2 = 0.76$—and group—$F(2,141) = 41.4$, $p < .001$, $\eta_p^2 = 0.37$. Post hoc analyses with Bonferroni corrections for multiple comparisons were conducted. As before, for the main effect of processing condition (collapsed across group and set size), recall in the alphabet processing condition was significantly worse than recall in the no processing condition (by 5.9%). In terms of the main effect of set size (collapsed across processing task and group), recall scores among all set sizes were significantly different (ranging from 4.5 to 28.7%), with recall being best for set size of 2 and progressively worsening through set size 6. For the main effect of group (collapsed across processing condition and set size), recall scores were significantly different among the three groups; the OHL listeners had significantly poorer recall compared with the ONH (by 6.8%) and the YNH listeners (by 14.7%) and the ONH listeners had poorer recall than the YNH listeners (by 7.9%). There was a significant two-way interaction of set size and group (collapsed across processing condition), $F(5.6,396.3) = 18.1$, $p < .001$, $\eta_p^2 = 0.21$. This interaction is illustrated in Figure 4 (bottom panel). As seen in the figure, all groups performed similar at the smallest set sizes and group differences become more apparent as set size increased. For comparison with recall, the mean word-recognition scores also are plotted (top panel) and do not show the same decline in performance with increasing set size. There also was a significant two-way interaction of set size and processing task, $F(3.9, 533.5) = 15.7$, $p < .001$, $\eta_p^2 = 0.10$, which is displayed in Figure 5 (corresponding recognition scores are also presented for reference). As illustrated, recall performance decreased with increasing set size, particularly with added processing with the alphabet judgment task. There was either no significant interaction between processing condition and group or a significant three-way interaction among processing task, set size, and group.

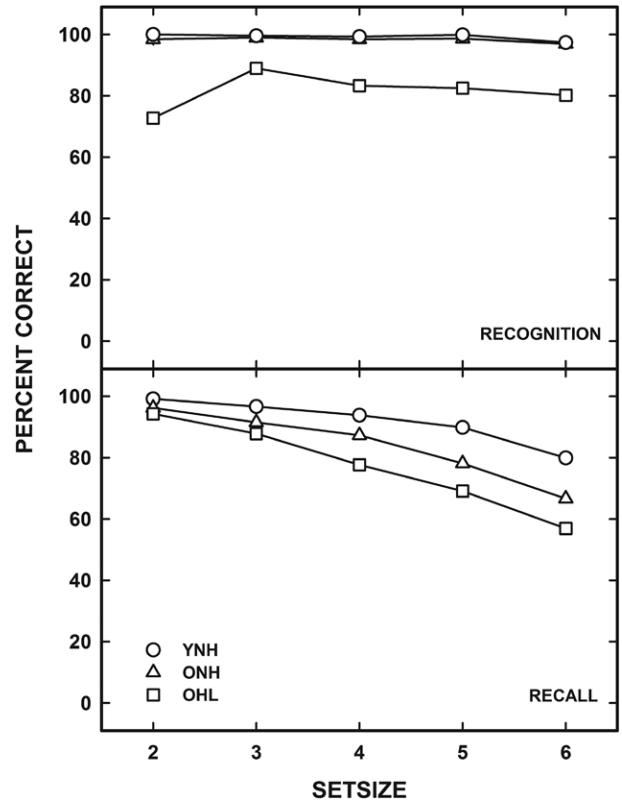

Fig. 4. The two-way interaction for recall (percent correct) between group and set size is shown (bottom panel). The corresponding data for recognition scores also are shown (top panel). OHL, older listeners with hearing loss; ONH, older listeners with normal hearing; YNH, younger listeners with normal hearing.

Span Scoring • The mean span scores obtained by the three listener groups in the two processing conditions are illustrated in Figure 6. The same pattern with the overall percent correct scoring procedure was observed with the span scoring procedure. All groups recalled more in the no processing condition than in the alphabet processing condition (also see Fig. 7). The YNH listeners recalled more than the two older listener groups, and the ONH listeners recalled more than the OHL group. A mixed model, repeated measures ANOVA was conducted and the results (with Greenhouse-Geiser corrections) revealed only significant main effects on recall for processing condition, $F(1.0,141.0) = 110.2$, $p < .001$, $\eta_p^2 = 0.44$, and group, $F(2,141) = 30.6$, $p < .001$, $\eta_p^2 = 0.30$. Again, these results

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

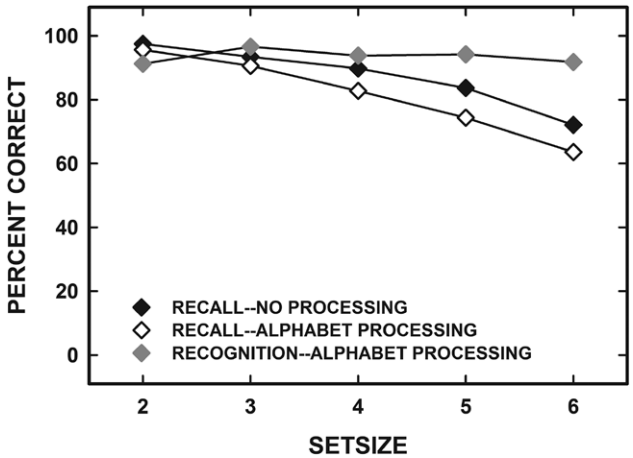

Fig. 5. The two-way interaction for recall (percent correct) between set size and processing task is shown (no processing = black diamonds; alphabet processing = open diamonds). For reference, the grey diamonds and line represent the corresponding recognition scores.

suggest that adding the alphabet judgment task to the WARRM test increased the cognitive processing load for all groups. This is reflected in the significantly poorer span scores in the alphabet processing condition compared with the no processing condition across all listener groups (0.8 difference between span scores in the processing conditions for the YNH and OHN listeners, and 0.5 for the OHL listeners). These results also demonstrate group differences, with YNH performing the best, followed by ONH, and the OHL listeners performing worst. Finally, these results confirm that scoring recall on the WARRM using either the percent correct or the span scoring procedures yields the same basic pattern of results (Fig. 8).

Correlations Between WARRM Span Results and Other Measures of Memory

Given that the two WARRM scoring procedures were strongly correlated (Fig. 8), we conducted further analyses with only one type of score. We selected the span score because a span score is unchanged whether the entire test is administered or the procedure is truncated (i.e., the test could be terminated once a span score is achieved, which could yield a more efficient procedure for clinical use). The WARRM recall results were expected to be related to other measures involving cognition and memory across the listener groups (see Table 3 for the mean performance on these memory/cognitive measures) and Pearson *r* correlations were performed to explore these relations. When collapsed across group, span in both processing conditions was significantly correlated with all memory/cognitive measures, as expected. Table 4 lists the correlations that ranged from fair to moderate. This correlation analysis was repeated for each listener group separately and those results are reported in Table 5.

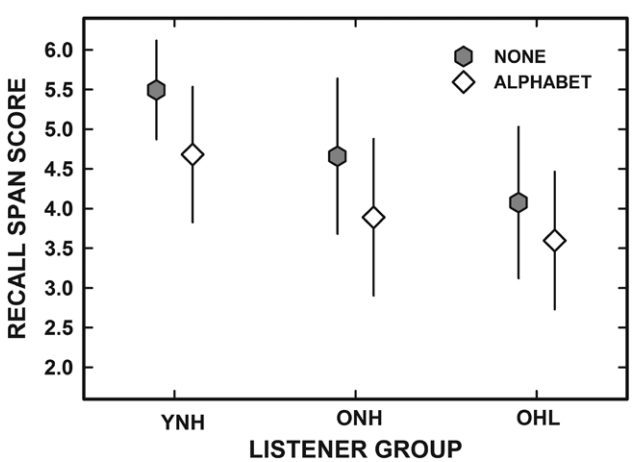

Fig. 6. The mean recall performance measured using the span scoring method is plotted as a function of listener group for the no processing condition (grey hexagons) and for the alphabet processing condition (open diamonds). The error bars represent 1 SD. OHL, older listeners with hearing loss; ONH, older listeners with normal hearing; YNH, younger listeners with normal hearing.

Association Between WARRM Span and 50%-Point Thresholds on the WIN#2

Mean WIN#2 performances, in dB S/N (1 SD), were 4.4 (1.2), 5.9 (1.6), and 13.7 (3.2) for the YNH, ONH, and OHL listeners, respectively. The association between WARRM span and 50%-point thresholds on the WIN#2 was examined with regression analyses to explore the amount of variance in speech in noise performance that was explained by auditory working memory. Because age and high-frequency hearing loss have been shown to be associated with WIN thresholds (Wilson 2011), age and the high-frequency pure-tone threshold average (HFPTA; bilateral average of 1000, 2000, and 4000 Hz) were used as control variables. Six separate hierarchical regression analyses were computed to examine the association for WIN#2 performance and the WARRM memory span for both processing conditions for each listener group. The control variables (age and HFPTA) were entered into the first step, followed by entry of WARRM span score on the second step. For all groups and with both processing conditions, the control variables explained a significant portion of variance in WIN#2 performance for all listener groups, but especially for the OHL listeners. High-frequency pure-tone average was the primary control variable that contributed to this variance (see Tables 4 and 5 in Supplemental Digital Content, http://links.lww.com/EANDH/A277, for regression results where only the control variable(s) were significant). After controlling for age and HFPTA, WARRM memory span results account for a significant portion (7.4%) of WIN#2 performance in the ONH listeners (alphabet processing condition) and for a significant portion (4.3%) of WIN#2 performance in the OHL listeners (no processing condition; Table 6).

DISCUSSION

The WARRM test was developed to augment the information gathered during the standard audiologic assessment by adding new information about auditory working memory for speech. Below, we will (1) interpret the important findings of the present study regarding the WARRM, (2) discuss the potential advantages of the WARRM over other working memory tests, and (3) suggest possible clinical applications of the WARRM in rehabilitative audiology.

Discussion of WARRM Results from the Present Study

As discussed, the present study provides a number of results that will be useful in guiding the further refinement of the

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

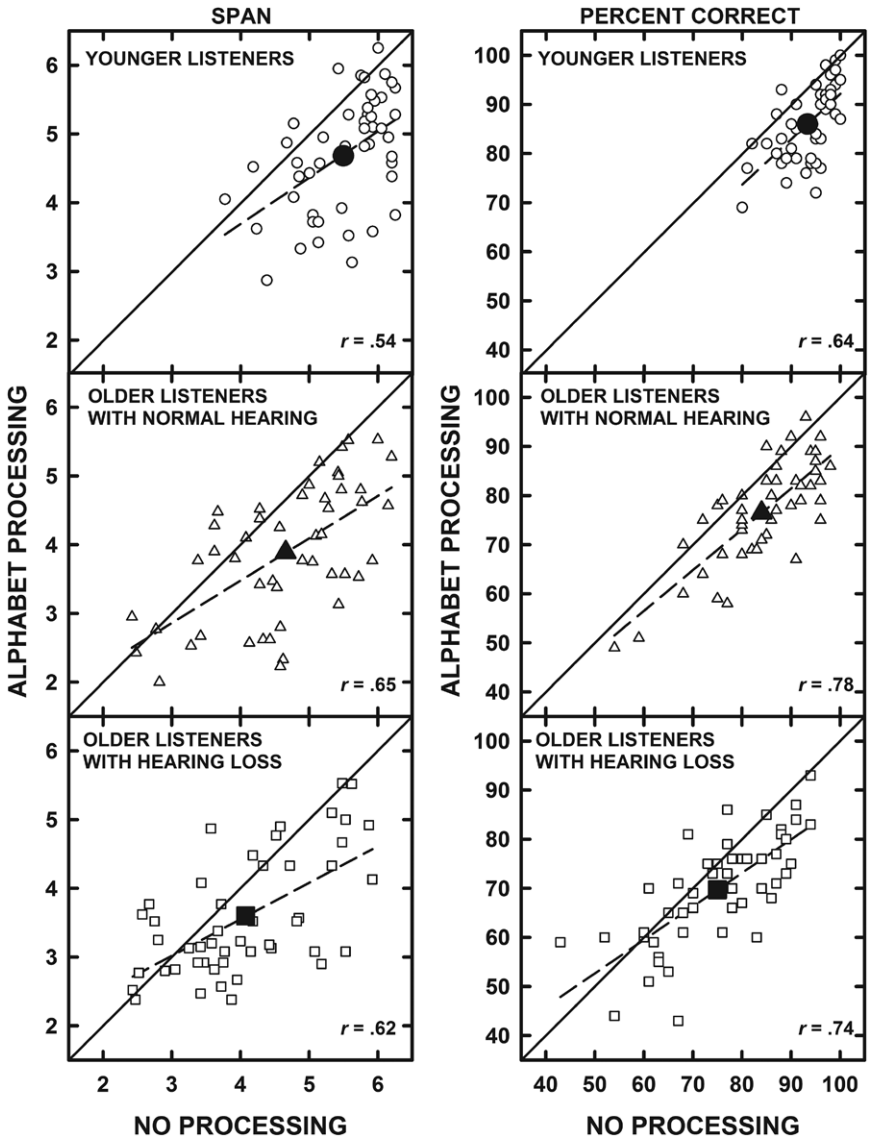

Fig. 7. Recall scores in the alphabet processing condition (ordinate) are plotted as a function of recall scores in the no processing condition (abscissa) for the span scoring method (first column) and the percent correct scoring method (second column). The results for the three groups are shown in separate panels along with the Pearson *r* correlations. The unfilled symbols represent individual datum points and the large, filled symbols represent the mean performance. The linear regression line for each group is plotted. The data in the figure were jittered using an additive algorithm from −0.25 to 0.25 in 0.05 steps to avoid superimposition.

WARRM test. These findings confirm the feasibility of combining testing of recognition and recall, demonstrate the value of incorporating cognitive processing demands when testing auditory recall, provide evidence of construct validity in terms of associations with results of memory measures and expected associations with speech in noise test results, and raise questions about generalizability of the findings to other populations.

Combining Recognition and Recall into One Test

In designing the WARRM, we used the monosyllabic word-recognition materials commonly used by audiologists and incorporated them into a standard working memory paradigm. The WARRM results provide two scores, a word-recognition score, and a recall score. Ideally, both the word-recognition and recall scores from the WARRM would yield clinically useful information without contamination by the integration of the test procedures. In terms of word recognition on the WARRM, as expected, the YNH and ONH listeners had performance at or near ceiling and the OHL listeners had poorer but still good word-recognition performance. In general, the word-recognition performance for the OHL listeners was stable across set sizes, except for set size two, probably because that set size contained a few words for which the OHL listeners obtained lower scores (see Wilson & McArdle 2015 who demonstrated that recognition performance among a set of words is not homogeneous for listeners with hearing loss; see also Table 3 in Supplemental Digital Content, http://links.lww.com/EANDH/A277, for an item analysis of each word). Word recognition in the alphabet processing condition was ~2% better than in the no processing condition. This difference is not considered to be consequential and is presumably due to learning/practice effects because the no processing condition was always given before the alphabet

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

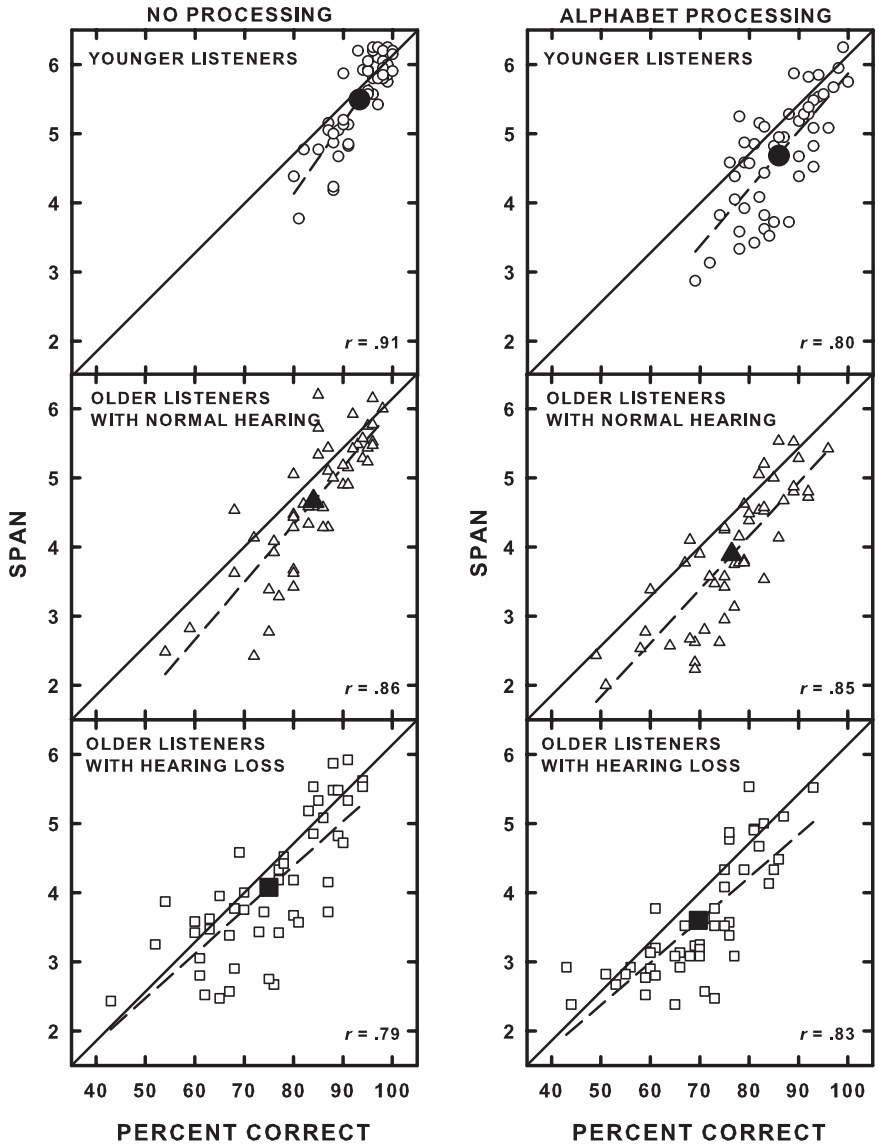

Fig. 8. Span scores (ordinate) are plotted as a function of percent correct recall scores (abscissa) for the no processing task condition (first column) and the alphabet processing task condition (second column). The results for the three groups are shown in separate panels along with the Pearson *r* correlations. The unfilled symbols represent individual datum points and the large, and the filled symbols represent the mean performance. The linear regression line for each group also is plotted. The data in the figure were jittered using an additive algorithm from −0.25 to 0.25 in 0.05 steps to avoid superimposition.

TABLE 3. The mean (and 1 SD) of the memory/cognitive variables measured in the three listener groups

	Younger Normals	Older "Normals"	Older Hearing Loss
Digit span			
Forward	9.3 (2.1)	8.9 (1.8)	8.3 (2.0)
Backward	8.3 (2.4)	7.8 (1.8)	7.2 (1.6)
Sequencing	10.0 (1.8)	8.9 (1.6)	8.1 (2.0)
Auditory free recall	8.2 (2.2)	6.6 (1.6)	4.9 (1.7)
Visual free recall	8.2 (2.1)	5.5 (2.1)	4.6 (2.0)
Vocabulary	34.9 (7.3)	37.9 (9.9)	32.6 (7.3)
Montreal cognitive assessment	28.2 (1.5)	26.5 (3.0)	26.1 (2.5)

processing condition. Importantly, word-recognition accuracy was not detrimentally affected by increasing set size for any listener group. Also noteworthy is that WARRM word-recognition scores were obtained binaurally, a more ecologically valid condition of everyday listening. Furthermore, the item-by-item recall score takes into account the word that was recognized. A binaural word-recognition score would not replace traditional ear-specific or diagnostic testing. The WARRM, however, could be administered monaurally if ear-specific information was of interest (e.g., Smith & Pichora-Fuller 2015). Future research would be needed to determine if WARRM recognition scores obtained monaurally (with the full or truncated procedures) are equivalent to and could possibly replace traditional word-recognition testing.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 4. The Pearson *r* correlation coefficients among the span scores (no processing and alphabet processing conditions) and the memory/cognitive measures collapsed across all listeners (n = 144)

	No Processing	Alphabet Processing
Digit span forward	0.29, $p < .001$	0.31, $p = .002$
Digit span backward	0.42, $p < .001$	0.32, $p < .001$
Digit span sequencing	0.31, $p < .001$	0.45, $p < .001$
Auditory free recall	0.46, $p < .001$	0.41, $p < .001$
Visual free recall	0.55, $p < .001$	0.51, $p < .001$
Vocabulary	0.54, $p < .001$	0.39, $p < .001$
Montreal cognitive assessment	0.51, $p < .001$	0.57, p < .001

As expected, despite stable word-recognition performance across set size, WARRM recall results declined with increasing set size. Based on this pattern of results, we would expect clinicians to be able to obtain information about general word-recognition abilities, while simultaneously obtaining new information about working memory for speech. Furthermore, for all three groups, there was good agreement between the results obtained using the percent correct recall scoring and the span scoring procedures, indicating that either procedure could be implemented in the clinic. If it was important to fix the length of the test for the purposes of word-recognition testing, then the percent correct scoring method could be used. If, however, shortening the memory test was a priority, then the span scoring approach would allow the examiner to terminate the test once the set size is reached where recall performance plateaus because better recall performance should not be achieved by the listener at larger set sizes.

Choice of Processing Condition

In the no processing condition, working memory resources were required only for hearing and repeating the target word, similar to the kind of cognitive load typically required in clinical tests of word recognition. Differences in recall between the groups in the no processing condition most likely reflect differences in the auditory demands of the recognition task and are likely to be associated with pure-tone threshold elevation and/or suprathreshold auditory processing deficits, that is, recall would be poorer for those with poorer hearing because increased auditory processing demands would deplete the working memory resources available for recall. In the alphabet condition, additional working memory resources are required because the listener has to think about the spelling of the word and make a decision about the place of the target word in the alphabet. The additional nonauditory processing demanded by the alphabet task was expected to result in reduced recall because it would pose an increased demand on working memory resources that would further deplete the spare capacity available for storage, that is, the alphabet processing condition would reflect a combination of auditory and nonauditory processing demands on resource allocation in working memory. Indeed, recall in the alphabet processing condition was significantly worse than recall in the no processing condition for all three groups of listeners. In effect, the score in the alphabet processing condition provides a new and more ecologically relevant measure of how well listeners can use information for other purposes once words have been heard, thereby more effectively revealing differences in groups of listeners who vary in age and degree of hearing loss. Furthermore, because this additional processing requirement differentiates WARRM from tests in which the task is simply to repeat words, we decided to focus on the administration of WARRM with the alphabet processing condition as we move forward with test refinement. Future variants of WARRM could incorporate a no processing condition or other types of processing conditions.

Construct Validity

Construct validity of the WARRM was demonstrated by the overall finding that WARRM recall was significantly correlated with other memory and/or cognitive measures across participants. In another study, Smith and Pichora-Fuller (2015) showed that WARRM span (alphabet processing) was significantly correlated with reading span for OHL ($r = .55$), further demonstrating construct validity of the WARRM. Nevertheless, the correlations between WARRM and the memory/cognitive measures in the present study (Table 5) were not strong (i.e., they were not >.80; Franzblau 1958). Taken together, these correlations suggests that WARRM recall, in either processing condition, assesses abilities in common with other memory measures, as well as assessing unique memory abilities, which also may explain why correlations between WARRM span and memory or cognitive measures differed in terms of number, strength, and significance at the group level. Within each listener group, the correlations between WARRM and other memory or cognitive measures were likely influenced by a combination of group differences in demographic characteristics, auditory processing abilities, task requirements, and test modality.

TABLE 5. Pearson *r* correlation coefficients (and *p* values) relating span (in each processing condition) to the memory/cognitive measures are listed for each listener group

	DSF	DSB	DSS	AFR	VFR	Vocabulary	MoCA
No processing							
YNH	.34 (.017)	ns	ns	.29 (.045)	.31 (.033)	.41 (.004)	.40 (.005)
ONH	.51 (.001)	.32 (.026)	.48 (.001)	.44 (.002)	ns	.53 (<.001)	.55 (<.001)
OHL	.29 (.046)	ns	ns	Ns	.41 (.004)	ns	.34 (.018)
Alphabet processing							
YNH	ns	ns	ns	Ns	ns	.49 (<.001)	.35 (.015)
ONH	.38 (.008)	.48 (.001)	.49 (<.001)	Ns	.33 (.024)	.52 (<.001)	.67 (<.001)
OHL	ns	ns	ns	Ns	.42 (.003)	ns	.37 (.010)

AFR, auditory free recall; DSB, digit span backward; DSF, digit span forward; DSS, digit span sequencing; MoCA, Montreal cognitive assessment; ns, not significant; ONH, older listeners with normal hearing; OHL, older listeners with hearing loss; VFR, visual free recall; YNH, younger listeners with normal hearing.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 6. Hierarchical regression results predicting Words-in-Noise (#2) performance from WARRM span scores in the older listeners

	R	$R^2\Delta$	*B**	*β*†	Partial Correlation	*F*/*t*‡	*p*
Older listeners with normal hearing							
Step 1: control variables	.543	.295				9.40	<.001
Age			0.072	0.237	0.229	1.83	.074
Hearing loss (HFPTA)			0.119	0.429	0.414	3.30	.002
Step 2: WARRM	.607	.074				5.13	.028
Span score (alphabet processing)			−0.470	−0.297	−0.271	−2.27	.028
Older listeners with hearing loss							
Step 1: control variables	0.761	0.579				30.92	<.001
Age			0.079	0.142	0.139	1.44	.157
Hearing loss (HFPTA)			0.250	0.720	0.705	7.29	<.001
Step 2: WARRM	0.788	0.043				4.98	.031
Span score (no processing)			−0.699	−0.207	−0.207	−2.23	.031

*Unstandardized regression coefficient from that step.
†Standardized regression coefficient from that step.
‡*F* corresponding to R^2 change; *t* corresponding to *β*.
WARRM, word auditory recognition and recall measure.

Association Between Memory and Speech in Noise Performance

Significant associations between cognitive or working memory measures and speech in noise performance have been found in many studies. The WIN test is an example of a well-documented instrument used to assess word-recognition performance in noise. We assessed the correlations between the WARRM span and WIN#2 50%-point thresholds for all listeners and in each of the three listener groups. Having controlled for age and hearing loss (Wilson 2011), we found that WARRM memory span is significantly, but weakly, related to WIN#2 threshold for both groups of older listeners, but not for younger listeners. For young listeners, regression analyses showed that WARRM memory span was not associated with WIN#2 performance, presumably because the variance in their measures was too restricted. Even with the additional alphabet processing task, the WIN#2 task presumably imposed only minimal processing demands on the YNH group. The results of the hierarchical regression analyses showed that the WARRM span explained an additional, but small, proportion of variance above age and hearing loss, but only in the alphabet processing condition for the ONH listeners (7.4%) and the no processing condition for the OHL listeners (4.3%). These results are consistent with the interpretation that the OHL listeners found listening to be sufficiently demanding to affect recall even in the no processing condition, whereas the ONH group listeners did not find it so demanding until the additional alphabet processing task was added.

Humes (2007) suggested that sufficient audibility of speech is achieved when the signal is 15 dB sensation level (re: each pure-tone threshold). In the present study, we used a presentation level of 70 or 80 dB HL (depending on the pure-tone average of the participant), but it is possible that some high-frequency information was not completely audible for the listeners with hearing loss (Fig. 1; also see Smith et al. 2012). If we had shaped the presentation level of the stimuli according to the hearing loss of the participant as suggested by Humes, then WARRM span in the alphabet processing condition may have emerged as a significant predictor in WIN#2 performance for the OHL listeners. We did not do this in the present study because most clinicians present speech stimuli to patients at an absolute presentation level via CD routed through an audiometer. The proportion of variance in WIN#2 performance explained by WARRM memory span for both older groups, however, is in line with other reports that cognition explains a small but significant percent of the variance (5 to 15%) in speech understanding (e.g., Akeroyd 2008; Houtgast & Festen 2008).

Generalizability

Of course, the results of this study are limited to the samples in the present study and may not be generalizable to other samples. For example, most participants in the OHL group were from the Veteran population and predominately male, whereas the participants in the YNH and ONH groups were recruited from the community and included more females. Although such gender differences may affect some aspects of hearing, they are not likely to affect working memory results significantly (e.g., Redick et al. 2012). Nevertheless, it is possible that differences among the groups in demographic characteristics, gender, and other variables contributed to group differences on the WARRM; however, all groups met the eligibility requirements for the study (e.g., no comorbid conditions, passing the MMSE, etc.) and were similar in all other nonauditory criteria we evaluated. Even if the ONH and OHL groups were matched on nonhearing variables, we would still expect differences in the groups due to differences in auditory processing on the WARRM. The WARRM was administered in the auditory domain as an attempt to quantify the interplay between memory processing and auditory processing within and between people. We anticipate that listeners with hearing loss (e.g., OHL) would allocate more working memory resources to processing auditory stimuli than listeners with good hearing (e.g., ONH) such that group differences on the WARRM would be expected even if all other variables were similar between the two older groups. Given the diversity of patient characteristics in clinical populations, future studies with the WARRM will need to compare groups that are matched or that contrast on demographic variables (e.g., gender, etc.) so that the influence of these characteristics on test results can be ruled out or appropriately taken into account when test results are interpreted clinically.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Advantages of WARRM Over Other Working Memory Measures

Audiologists have become increasingly interested in measuring working memory in the clinic because working memory may play an important role in speech understanding, especially in complex listening situations. Measuring it may be a useful addition to the rehabilitative audiologic workup, (e.g., Pichora-Fuller & Singh 2006; Souza 2012). To date, most studies have used the reading span working memory measure, especially when the purpose of the research has been to examine interindividual differences in working memory capacity. For a number of reasons, the reading working memory span test may not be the most clinically feasible test in audiology clinics. A listening working memory span test such as the WARRM may be a more feasible and relevant measurement. First, the test administration time for the typical reading working memory test is approximately 20 to 30 min and would occupy a significant portion of valuable and limited appointment time in most clinics, whereas, the WARRM takes approximately 10 min to administer. Second, the reading span test is administered in the visual modality. As previously discussed, measuring working memory using speech stimuli in the auditory domain, the same domain in which individuals understand speech, may be the more ecologically valid if the goal is to assess the interplay between auditory and cognitive processing experienced by an individual in everyday listening situations (Baldwin & Ash 2011; Besser et al. 2013). Indeed, there is evidence to suggest that working memory is domain and modality specific (e.g., Pichora-Fuller et al. 1995; see Besser et al. 2013 for review). Clearly, it would not be possible to use the reading span test to evaluate the effects on an individual of varying processing demands by manipulating auditory demands, (e.g., masking conditions). Another practical consideration is the possible effect of vision loss or low literacy level on reading working memory span. Third, measuring intraindividual differences in working memory span under various listening conditions has potential as an outcome measure in audiologic rehabilitation insofar as any improvements in auditory functioning would more likely be detected by a test of listening than in a reading working memory span (Pichora-Fuller 2006a, b, 2008; Pichora-Fuller & Singh 2006). Future studies will investigate the use of the WARRM as a metric to assess intraindividual differences.

We developed the WARRM to be a domain- and modality-specific test of auditory working memory for speech that would overcome the limitations of the reading span working memory measure in assessing listening. The WARRM incorporates audio-recorded speech materials from established clinical word-recognition tests into a well-established working memory span test procedure. The measure is presented via CD through the audiometer so that the presentation level can be adjusted according to the pure-tone hearing thresholds of the listener following established audiologic testing methods. The WARRM takes approximately 10 min to administer in its entirety (all 100 words), including practice. If the span scoring method is used rather than the percent correct method then testing time could be shortened by 50% or more depending on the individual span score because lower span scores would require less test administration time (Pichora-Fuller 2003, 2006a, b, 2007). Terminating the WARRM when a span score is achieved may prove to be more clinically feasible for busy clinicians and potentially improve acceptability of the WARRM by patients who may find recall with higher set sizes frustrating. Because the WARRM is administered in the auditory modality, it could be easily adapted to assess intraindividual differences in working memory under different listening conditions and also serve as an outcome measure (e.g., to evaluate the benefit of hearing aid use). Finally, the WARRM may be a useful counseling tool for patients as the results may provide insight into patient complaints that speech understanding is effortful or tiring (e.g., in demanding listening situations, a lower span would suggests that the individual may experience higher listening effort, while a higher span would suggests that the individual may experience lower listening effort), even when both individuals have good word-recognition scores.

SUMMARY AND CONCLUSIONS

Experimental research suggests that interindividual differences in working memory may be a key indicator of who will achieve improvements in speech in noise performance when using certain hearing aid processing features, such as compression, noise reduction, or frequency compression (e.g., Lunner 2003; Gatehouse et al. 2006a, b; Foo et al. 2007; Moore 2008; Lunner et al. 2009; Rudner et al. 2011a; Arehart et al. 2013; Ng et al. 2013, 2014; Desjardins & Doherty 2014; Souza & Sirow 2014). Clinicians, therefore, may alter the way in which hearing aids are programmed if the working memory capacity of individual patients could be determined by a clinically feasible measure of working memory, such as the WARRM. Furthermore, listeners with lower working memory may be identified as needing additional audiologic rehabilitation, above and beyond hearing aids, to reduce the perceptual effort of understanding speech. For example, listeners could use compensatory top-down cognitive strategies by using supportive context to facilitate speech understanding in difficult listening situations (e.g., Pichora-Fuller 2003, 2006a, b, 2008; Pichora-Fuller & Singh 2006, for a discussion; see also Pichora-Fuller et al. 1995; Kricos 2006; Sheldon et al. 2008), and current research examining alternative approaches for hearing-aid users are ongoing (e.g., Ferguson & Henshaw 2015a, b).

Intraindividual differences in working memory may offer a new outcome domain in audiologic rehabilitation. A goal for audiologic intervention would be to reduce the auditory processing demands involved in speech understanding so that fewer working memory resources are allocated to listening, thereby preserving more working memory capacity for storing the information that has been heard or for using for more elaborate processing. Therefore, working memory span could be assessed before and after an auditory intervention to determine how effective the intervention was in reducing the sensory processing load during speech understanding.

In conclusion, the WARRM was designed to fill a gap by providing a clinically feasible, auditory working memory measure that could be used by clinicians to plan and evaluate audiologic rehabilitation. Before implementation, further research is needed to establish the psychometric properties of the measure, including the test–retest reliability, and to develop normative data for the measure. The results of the WARRM from patient populations that differed in various demographic characteristics

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

should be explored further to assess the generalizability of the measure. The use of the WARRM as a tool to examine intraindividual differences in working memory under different listening conditions, such as comparing aided and unaided conditions, or presenting the WARRM in quiet compared with noisy conditions, also should be evaluated. In addition, future research should establish an evidence base for the clinical utility of the unaided WARRM in terms of determining candidacy for hearing aid processing options and for planning and evaluating other rehabilitation interventions.

ACKNOWLEDGMENTS

Beth Bailey is acknowledged for providing statistical support. Richard H. Wilson is acknowledged for the construction and production of the stimulus materials.

Supported by the Department of Veterans Affairs (VA), Veterans Health Administration, Office of Research and Development, Rehabilitation Research and Development (RR&D) Service, Washington DC through a Career Development Award (#C6394W) to the first author and by the RR&D Auditory Vestibular Research Enhancement Award Program (REAP; #C4339F) to the first and third authors.

The first two authors made substantial contributions to conception and design of the work. G.A. acquired the data from the participants. S.L.S. performed the data analysis with consultation from a statistician. All the authors contributed to the interpretation of data. S.L.S. drafted the article and the other authors contributed substantially to revising the article critically for important intellectual content. All authors provided final approval of the version to be published.

The contents of this manuscript do not represent the views of the Department of Veterans Affairs or the United States government.

The authors declare no other conflicts of interest.

Address for correspondence: Sherri L. Smith, Audiologic Rehabilitation Laboratory, Auditory Vestibular Research Enhancement Award Program (AVREAP), Audiology 126, Mountain Home, TN 37684, USA. E-mail: sherri.smith@va.gov

Received March 10, 2015; accepted April 4, 2016.

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. *Int J Audiol*, *47*(suppl 2), S53–S71.

Amichetti, N. M., Stanley, R. S., White, A. G., et al. (2013). Monitoring the capacity of working memory: Executive control and effects of listening effort. *Mem Cognit*, *41*, 839–849.

American National Standards Institute. (2010). *American National Standard Specification for Audiometers (ANSI S3.6-2010)*. New York.

Arehart, K. H., Souza, P., Baca, R., et al. (2013). Working memory, age, and hearing loss: Susceptibility to hearing aid distortion. *Ear Hear*, *34*, 251–260.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychol Learn Motiv*, *8*, 47–89.

Baddeley,A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 28–61). New York, NY: Cambridge University Press.

Baldwin, C. L., & Ash, I. K. (2011). Impact of sensory acuity on auditory working memory span in young and older adults. *Psychol Aging*, *26*, 85–91.

Besser, J., Koelewijn, T., Zekveld, A. A., et al. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise–a review. *Trends Amplif*, *17*, 75–93.

Bilger, R. C. (1984). Speech recognition test development. In: E. Elkins (Ed.), *Speech Recognition by the Hearing Impaired. ASHA Reports 14*. Rockville, MD: American Speech-Language-Hearing Association.

Broadbent, D. E. (1958). Effect of noise on an "intellectual" task. *J Acoust Soc Am*, *30*, 824–827.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Otolaryngol*, *40*, 62–71.

Cocchini, G., Logie, R. H., Della Sala, S., et al. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Mem Cognit*, *30*, 1086–1095.

Crum, R. M., Anthony, J. C., Bassett, S. S., et al. (1993). Population-based norms for the mini-mental state examination by age and educational level. *JAMA*, *269*, 2386–2391.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *J Verbal Learning Verbal Behav*, *19*, 450–466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *J Exp Psychol Learn Mem Cogn*, *9*, 561–584.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychon Bull Rev*, *3*, 422–433.

Department of Veterans Affairs. (2006). *Speech recognition and identification materials*. *Disc 4.0*. Department of Veterans Affairs.

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear*, *35*, 600–610.

Dillon, H., Cameron, S., Tomlin, D., et al. (2014). Comments on "Factors influencing tests of auditory processing: A perspective on current issues and relevant concerns" by Tony Cacace and Dennis McFarland. *J Am Acad Audiol*, *25*, 699–703.

Doherty, K. A., & Desjardins, J. L. (2015). The benefit of amplification on auditory working memory function in middle-aged and young-older hearing impaired adults. *Front Psychol*, *6*, 721.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope*, *58*, 955–991.

Engle, R. W., Cantor, J., Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *J Exp Psychol Learn Mem Cogn*, *18*, 972–992.

Engle, R. W., Kane, M. J., Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 102–134). Cambridge: Cambridge University Press.

Ferguson, M. A., & Henshaw, H. (2015a). Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Front Psychol*, *6*, 556.

Ferguson, M., & Henshaw, H. (2015b). How does auditory training work? Joined-up thinking and listening. *Semin Hear*, *36*, 237–249.

Finney, D. J. (1952). *Statistical Method in Biological Assay*. London: C. Griffen.

Folstein, M. F., Folstein, S. E., McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, *12*, 189–198.

Foo, C., Rudner, M., Rönnberg, J., et al. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J Am Acad Audiol*, *18*, 618–631.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Atten Percept Psychophys*, *71*, 1360–1374.

Franzblau, A. N. (1958). *A Primer of Statistics for Non-Statisticians*. New York, NY: Harcourt, Brace and World.

Gatehouse, S., Naylor, G., Elberling, C. (2006a). Linear and nonlinear hearing aid fittings–1. Patterns of benefit. *Int J Audiol*, *45*, 130–152.

Gatehouse, S., Naylor, G., Elberling, C. (2006b). Linear and nonlinear hearing aid fittings–2. Patterns of candidature. *Int J Audiol*, *45*, 153–171.

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *J Speech Lang Hear Res*, *40*, 423–431.

Hirsh, I. J., Davis, H., Silverman, S. R., et al. (1952). Development of materials for speech audiometry. *J Speech Hear Disord*, *17*, 321–337.

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. *Int J Audiol*, *47*, 287–295.

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol*, *18*, 590–603.

Humes, L. E., & Floyd, S. S. (2005). Measures of working memory, sequence learning, and speech recognition in the elderly. *J Speech Lang Hear Res*, *48*, 224–235.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Humes, L. E., Kidd, G. R., Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front Syst Neurosci*, *7*, 55.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychol Rev*, *99*, 122–149.

Kiessling, J., Pichora-Fuller, M.K., Gatehouse, S., et al. (2003). Candidature for and delivery of audiological services: Special needs of older people. *Int J Audiol*, *42*(suppl 2), S92–S101.

Kreul, E. J., Bell, D. W., Nixon, J. C. (1969). Factors affecting speech discrimination test difficulty. *J Speech Hear Res*, *12*, 281–287.

Kricos, P. B. (2006). Audiologic management of older adults with hearing loss and compromised cognitive/psychoacoustic auditory processing capabilities. *Trends Amplif*, *10*, 1–28.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear*, *19*, 1–36.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int J Audiol*, *42*(suppl 1), S49–S58.

Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *J Am Acad Audiol*, *18*, 604–617.

Lunner, T., Rudner, M., Rönnberg, J. (2009). Cognition and hearing aids. *Scand J Psychol*, *50*, 395–403.

McArdle, R., & Wilson, R. H. (2008). Predicting word-recognition performance in noise by young listeners with normal hearing using acoustic, phonetic, and lexical variables. *J Am Acad Audiol*, *19*, 507–518.

McFarland, D. J., & Cacace, A. T. (2014). Modality specificity is the preferred method for diagnosing the auditory processing disorder (APD): Response to Moore and Ferguson. *J Am Acad Audiol*, *25*, 698–699.

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J Assoc Res Otolaryngol*, *9*, 399–406.

Moore, D. R., & Ferguson, M. A. (2014). It is neither necessary nor desirable to test for abnormalities in other modalities when diagnosing auditory processing disorder (APD). *J Am Acad Audiol*, *25*, 697–698.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., et al. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*, 695–699.

Ng, E. H., Rudner, M., Lunner, T., et al. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int J Audiol*, *52*, 433–441.

Ng, E. H., Classon, E., Larsby, B., et al. (2014). Dynamic relation between working memory capacity and speech recognition in noise during the first 6 months of hearing aid use. *Trends Hear*, *18*.

Ohlman, H. M. (1959). Subject-word letter frequencies with applications to superimposed coding. In Proceedings of the International Conference on Scientific Information.

Park, D. C., Smith, A. D., Lautenschlager, G., et al. (1996). Mediators of long-term memory performance across the life span. *Psychol Aging*, *11*, 621–637.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *Int J Audiol*, *42*, S59–S67.

Pichora-Fuller, M. K. (2006a). Effects of auditory and cognitive aging on communication. *ASHA Perspect Hear Hear Disord Res Res Diagn*, *10*, 10–14.

Pichora-Fuller, M. K. (2006b). Perceptual effort and apparent cognitive decline: Implications for audiologic rehabilitation. *Sem Hear*, *27*, 284–293.

Pichora-Fuller, M. K. (2007). Audition and cognition: What audiologists need to know about listening. In C. Palmer, & R. Seewald (Eds.), *Hearing Care for Adults* (pp. 71–85). Stäfa, Switzerland: Phonak.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *Int J Audiol*, *47*(suppl 2), S72–S82.

Pichora-Fuller, M.K. (2013). Auditory and cognitive processing in audiologic rehabilitation. In J. Spitzer & J. Montano (Eds.), *Adult Audiologic Rehabilitation: Advanced Practices* (2nd ed.) (pp. 519–536). San Diego, CA: Plural Publishing.

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif*, *10*, 29–59.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*, *97*, 593–608.

Piquado, T., Benichov, J. I., Brownell, H., et al. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *Int J Audiol*, *51*, 576–583.

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory.
Q J Exp Psychol, *20*, 241–248.

Rabbitt, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngol Suppl*, *476*, 167–175; discussion 176.

Redick, T.S., Broadway, J.M., Meier, M.E., et al. (2012). Measuring working memory capacity with automated complex span tasks. *Euro J Psychol Assessment*, *28*, 164–171.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int J Audiol*, *42*(suppl 1), S68–S76.

Rönnberg, J., Arlinger, S., Lyxell, B., et al. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *J Speech Hear Res*, *32*, 725–735.

Rönnberg, J., Rudner, M., Foo, C., et al. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol*, *47*(suppl 2), S99–105.

Rönnberg, J., Rudner, M., Lunner, T., et al. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise Health*, *12*, 263–269.

Rönnberg, N., Stenfelt, S., Rudner, M. (2011). Testing listening effort for speech comprehension using the individuals' cognitive spare capacity. *Audiol Res*, *1*, e22.

Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7*, 31.

Rönnberg, N., Rudner, M., Lunner, T., et al. (2014). Memory performance on the auditory inference span test is independent of background noise type for young adults with normal hearing at high speech intelligibility. *Front Psychol*, *5*, 1490.

Rudner, M., Fransson, P., Ingvar, M., et al. (2007). Neural representation of binding lexical signs and words in the episodic buffer of working memory. *Neuropsychologia*, *45*, 2258–2276.

Rudner, M., Rönnberg, J., Lunner, T. (2011a). Working memory supports listening in noise for persons with hearing impairment. *J Am Acad Audiol*, *22*, 156–167.

Rudner, M. Ng, E. H. N., Rönnberg, N., et al. (2011b). Cognitive spare capacity as a measure of listening effort. *J Hearing Sci*, *11*, 47–49.

Schneider, B. A., Daneman, M., Pichora-Fuller, M. K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Can J Exp Psychol*, *56*, 139–152.

Shah, P., & Miyake, A. (1999). An introduction. In A Miyake & P Shah (Eds.), *Models of Working Memory* (pp. 1–29). New York, NY: Cambridge University Press.

Sheldon, S., Pichora-Fuller, M. K., Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J Acoust Soc Am*, *123*, 489–499.

Smith, S. L., & Pichora-Fuller, M. K. (2015). Associations between speech understanding and auditory and visual tests of verbal working memory: Effects of linguistic complexity, task, age, and hearing loss. *Front Psychol*, *6*, 1394.

Smith, S. L., Pichora-Fuller, M. K., Wilson, R. H., et al. (2012). Word recognition for temporally and spectrally distorted materials: the effects of age and hearing loss. *Ear Hear*, *33*, 349–366.

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychol Aging*, *11*, 333–341.

Souza, P. E. (2012). 20Q: Cognition measures-they might change the way you fit hearing aids. *Audiology Online*. Retrieved December 5, 2014, from http://www.audiologyonline.com/articles/august-20q-by-pamela-souza-6925.

Souza, P., & Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *Int J Audiol*, *54*, 705–713.

Souza, P. E., & Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing AIDS. *Am J Audiol*, *23*, 394–401.

Stewart, R., & Wingfield, A. (2009). Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. *J Am Acad Audiol*, *20*, 147–154.

Tillman, T. W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Auditory Test No. 6. Brooks Air Force Base, TX USAF School of Aerospace Medicine Technical Report.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York, NY: Psychological Corporation.

Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale-Revised: Manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV).* San Antonio, TX: NCS Pearson.

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J Am Acad Audiol*, *14*, 453–470.

Wilson, R. H. (2011). Clinical experience with the words-in-noise test on 3430 veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *J Am Acad Audiol*, *22*, 405–423.

Wilson, R. H., & McArdle, R. (2015). The homogeneity with respect to intelligibility of recorded word-recognition materials. *J Am Acad Audiol*, *26*, 331–345.

Wilson, R. H., Abrams, H. B., Pillion, A. L. (2003). A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. *J Rehabil Res Dev*, *40*, 321–327.

Wilson, R. H., McArdle, R., Roberts, H. (2008). A comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. *J Am Acad Audiol*, *19*, 496–506.

Wingfield, A., & Stine-Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of Aging and Cognition* (2nd ed) (pp. 359–416). Mahwah, NJ: Erlbaum.

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *J Am Acad Audiol*, *18*, 548–558.

Wingfield, A., McCoy, S. L., Peelle, J. E., et al. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J Am Acad Audiol*, *17*, 487–497.

Woodcock, R. W., McGrew, K. S., Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. Itasca, IL: Riverside Publishing.

Copyright © 2016 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.