

The Effect of Varying Test Administration and Scoring Procedures on Three Tests of (Central) Auditory Processing Disorder

DOI: 10.3766/jaaa.17063

Maria E. Pomponio*†
Stephanie Nagle*‡
Jennifer L. Smart*
Shannon Palmer§

Abstract

Background: There is currently no widely accepted objective method used to identify (central) auditory processing disorder ([C]APD). Audiologists often rely on behavioral test methods to diagnose (C)APD, which can be highly subjective. This is problematic in light of relevant literature that has reported a lack of adequate graduate-level preparation related to (C)APD. This is further complicated when exacerbated by the use of inconsistent test procedures from those used to standardize tests of (C)APD, resulting in higher test variability. The consequences of modifying test administration and scoring methods for tests of (C)APD are not currently documented in the literature.

Purpose: This study aims to examine the effect of varying test administration and scoring procedures from those used to standardize tests of (C)APD on test outcome.

Research Design: This study used a repeated-measures design in which all participants were evaluated in all test conditions. The effects of varying the number of test items administered and the use of repetitions of missed test items on the test outcome score were assessed for the frequency patterns test (FPT), competing sentences test (CST), and the low-pass filtered speech test (LPFST). For the CST only, two scoring methods were used (a strict and a lax criterion) to determine whether or not scoring method affected test outcome.

Study Sample: Thirty-three native English-speaking adults served as participants. All participants had normal hearing (as defined by thresholds of 25-dB HL or better) at all octave band frequencies from 500 to 4000 Hz, with thresholds of 55-dB HL or better at 8000 Hz. All participants had normal cognitive function as assessed by the Mini-Mental State Examination.

Data Collection and Analysis: Paired samples *t*-tests were used to evaluate the differences in test outcome when varying the CST scoring method. A $3 \times 2 \times 2$ repeated-measures factorial analysis of variance (ANOVA) was used to determine the effects of test, length, and repetitions on outcome score for all three tests of auditory processing ability. Individual 2×2 repeated-measures two-way ANOVAs were subsequently conducted for each test to further evaluate interactions.

Results: There was no effect of scoring method on the CST outcome. There was a significant main effect of repetition use for the FPT and LPFST, in that test scores were greater when corrected for repetitions. An interaction between test length and repetitions was found for the LPFST only, such that there was a greater effect of repetition use when a shorter test was administered compared with a longer test.

Conclusions: Test outcome may be affected when test administration procedures are varied from those used to standardize the test, lending itself to the broader possibility that the overall diagnosis of (C)APD may be subsequently affected.

*Towson University, Towson, MD; †Weill Cornell Medicine, New York, NY; ‡Connecticut Children's Medical Center, Hartford, CT; §Central Michigan University, Mount Pleasant, MI

Corresponding author: Maria E. Pomponio, Weill Cornell Medicine Hearing and Speech Center, New York, NY 10021; Email: map9385@med.cornell.edu

Portions of this paper were presented as a poster session at the 2015 American Academy of Audiology AudiologyNOW! Conference in San Antonio, TX; the 2016 American Auditory Society Conference in Scottsdale, AZ; and the 2016 American Speech-Language-Hearing Association Convention in Philadelphia, PA.

Key Words: (central) auditory processing, (central) auditory processing disorder, reliability and validity, scoring methods

Abbreviations: ANOVA = analysis of variance; (C)APD = (central) auditory processing disorder; CST = competing sentences test; FPT = frequency patterns test; LPFST = low-pass filtered speech test; SD = standard deviation

INTRODUCTION

The term (central) auditory processing disorder [(C)APD] is a term used to describe listening difficulties that result from the inability of an individual's central auditory nervous system to appropriately integrate auditory input (ASHA, 2005). (C)APD can manifest in a number of ways, occurring either as a stand-alone diagnosis or in the presence of another disorder (ASHA, 2005). According to the AAA, (C)APD is evaluated using a test battery approach and is diagnosed after an individual scores two or more standard deviations (SD) below the mean on two or more tests of auditory processing ability (AAA, 2010). It is estimated that (C)APD affects 2–3% of the pediatric population and anywhere from 23% to as much as 76% of adults; however, the prevalence of (C)APD is highly contested (Stach et al, 1990; Golding et al, 2004; O'Beirne et al, 2012). Specifically, there is currently a wealth of controversy surrounding the assessment and diagnosis of (C)APD, leading some audiologists to deny that (C)APD is a true clinical disorder (Sykes et al, 1997; Chermak et al, 1998; Emanuel, 2002; Bellis, 2006; Chermak et al, 2007; Dawes and Bishop, 2009).

There are several factors contributing to the controversy associated with (C)APD. Of primary significance is the fact that presently, there is no “gold standard” by which (C)APD is diagnosed nor is there any agreed-upon standard regarding the inclusion of objective measures to confirm or deny the presence of the disorder (Bellis, 2006; AAA, 2010; Emanuel et al, 2011). AAA (2010), ASHA (2005), and BSA (2011) continue to promote the use of a patient-specific test battery composed of sensitive, specific, and efficient behavioral tests to evaluate for (C)APD. Although AAA (2010) acknowledged situations in which the use of objective auditory evoked responses may be helpful in the assessment of (C)APD, issues with cost, equipment availability, and the length of these procedures have appeared to hinder the widespread acceptance of such methods (Katz et al, 2002). Resultantly, a diagnosis of (C)APD is often based solely on behavioral tests interpreted by audiologists, some of whom, admittedly, have little training in the area.

A 1998 survey by Chermak, Traynham, Seikel, and Musiek highlighted the daunting fact that most audiologists who test for (C)APD were <50% satisfied with the amount of preparation specific to (C)APD that they had during their graduate education. Eighty percent of the

179 respondents reported never taking a class explicitly dedicated to (C)APD. Furthermore, the mean number of clinical experience hours earned related to the assessment of (C)APD in children, adults, and the elderly was less than five hours (Chermak et al, 1998). A follow-up survey completed by Chermak et al (2007) confirmed the still-problematic, inadequate training specific to (C)APD earned during graduate education, citing less than nine hours of clinical experience related to (C)APD earned on campus and less than five hours earned off campus. Unfortunately, this lack of adequate exposure to (C)APD during the graduate program is not completely surprising, as a 1997 survey of 40 graduate programs conducted by Sykes, Tucker, and Herr confirmed that of five main areas of audiology (hearing aids, cochlear implants, tinnitus, vestibular disorders/dizziness, and [C]APD), (C)APD was viewed as the least important by faculty members. The present conundrum of potentially ill-prepared clinicians using behavioral measures to assess (C)APD is compounded by an additional acute factor aiding in the controversy; specifically, not all audiologists, regardless of their familiarity with (C)APD, may adhere to the guidelines for test administration and scoring procedures that were used to standardize tests of (C)APD, clinically or in research (McDermott et al, 2016). When audiologists do not follow these guidelines, not only are the validity and reliability of the test compromised but the clinical significance of a diagnosis of (C)APD is significantly diminished.

Bellis (2003) reported that two primary factors exist that can affect test reliability: patient and procedural variables. Patient variables are related specifically to the individual undergoing the assessment and include patient age, health, intelligence, language skills, and related disorders. Procedural variables are related to the test itself and include equipment and test calibration, practice, ceiling, and floor effects, as well as test administration and scoring variables (which may include the number of test items administered, repeating incorrectly answered test items to give the patient a second attempt to answer correctly, and using different methods to score the same test).

Currently, health care providers are under ever-increasing pressure to increase clinical productivity and decrease test time. Various studies and surveys have documented continued use of test procedures in audiology that are not best practice to shorten test time. The most notable example is the continued use

of monitored live voice and shortened word lists for word recognition testing (Wiley et al, 1995; Stewart, 2003; Schoepflin, 2012; Hormsby and Muller, 2013). Multiple studies also demonstrate researchers and clinicians using fewer test items than recommended on tests of auditory processing (Kelly, 2007; Tomlin et al, 2015; McDermott et al, 2016). In addition, many commercial auditory processing tests may come without sufficiently detailed testing information (Friberg and McNamara, 2010). Even when literature is available describing instructions for testing, instructions may be ambiguous. Specifically, instructions regarding repeating test items may be murky (“Pattern should not be repeated unless patient was not given a fair chance to hear the pattern”—Musiek, 1994; Musiek, 2002).

Ultimately, audiologists may vary procedural variables while administering tests of auditory processing ability because of a need for shortened test time, unclear instructions for a specific test, or a perception that the patient has not properly heard a test item. In addition, and as previously mentioned, studies have shown a lack of clinical education and preparation for this domain of audiology. Taken together, these factors may lead to audiologists administering clinical tests of auditory processing in nonstandard ways. It is not unreasonable to assume that modifying such factors may affect test outcome and reliability. When this occurs during (C)APD assessment, the integrity of a (C)APD diagnosis is subsequently diminished, fueling an ever-present controversy amid a lack of easily accessible and agreed-upon objective assessment methods.

METHODOLOGY

All research procedures used in this study were approved by the Institutional Review Board at Towson University. Thirty-seven adults (13 males and 24 females) ranging in age from 50 to 65 years ($M = 57.68$ years, $SD = 4.52$) were recruited to participate in this study. Participants were recruited using flyers around campus, Towson University faculty email databases, and word of mouth. All participants had unremarkable otologic and medical histories, and normal cognitive function as assessed by the Mini-Mental State Examination (Folstein et al, 1975).

Testing was completed at Towson University in a double-walled sound-treated test suite. All tests were administered by the first author as part of a thesis study. Each evaluation was concluded in one session, with breaks given to participants as needed. Tympanometry and pure-tone assessment were completed using a TympStar tympanometry bridge and a GSI-61 two-channel audiometer, respectively. All tests of (C)APD were administered via compact disc (CD). An external CD player routed through both channels of the GSI-61 audiometer was used to present all recorded

materials through Etymotic Research-3A insert earphones. Calibration was performed before each test so that the VU meter peaked at 0-dB HL for both channels of the audiometer using the calibration tone provided on the CD.

All participants had mobile tympanic membranes (0.2–1.8 mL) with a peak pressure no < -150 daPa. Participants who demonstrated normal hearing (as defined by thresholds of 25-dB HL or better at all octave band frequencies from 500 to 4000 Hz, and hearing thresholds of 55-dB HL or better at 8000 Hz to account for presbycusis; Cox et al, 2008) with no significant asymmetries between ears and no air-bone gaps were subsequently evaluated for (C)APD.

Tests of central auditory processing ability included the frequency pattern test (FPT; Musiek and Pinheiro, 1987), the competing sentences test (CST; Willeford and Burleigh, 1994), and the 1000-Hz, male speaker version of the low-pass filtered speech test (LPFST; Auditec Inc., St. Louis, MO). Adults were chosen as participants because of the higher prevalence of (C)APD found in adults (24–76% in adults aged 55–75 years; 70% in adults more than 60 years) compared with children (5%; Chermak, 2016). Tests were selected based on AAA (2010) and ASHA (2005) guidelines that recommend using a test battery that evaluates a variety of auditory processes including, but not limited to, temporal processing (e.g., FPT), dichotic listening (e.g., CST), and monaural low-redundancy speech perception (e.g., LPFST) when assessing for (C)APD. Furthermore, the FPT, CST, and LPFST have been documented as commonly used behavioral measures of (C)APD (Emanuel, 2002) with documented sensitivity, specificity, and efficiency values (Musiek et al, 2011).

Test and ear order were randomized for each participant. For all tests, each participant was first administered the standard number of test items. Following completion of all test items, any items answered incorrectly were repeated and marked separately. All participants underwent testing in all conditions. Each test was then scored by the first author in the following four conditions: (a) standard number of test items, not including repetitions; (b) standard number of test items, corrected for repetitions; (c) adapted (shortened) number of test items, not including repetitions; and (d) adapted (shortened) number of test items, corrected for repetitions. The adapted number of test items was a number less than the standard number administered for each test under the premise that, for time-saving purposes, clinicians are likely to use fewer test items rather than a greater number of test items when deviating from standard procedures, as noted in the studies discussed previously (Kelly, 2007; Tomlin et al, 2015; McDermott et al, 2016). For the CST only, two scoring methods were compared to determine if using a stricter criterion (Musiek and Pinheiro, 1985) would result in

any difference in the overall test score compared with using a more lax criterion (Willeford and Burleigh, 1994). Test-specific procedures are outlined in Table 1.

RESULTS

Of the 37 adults recruited to participate in this study, 4 of the 37 volunteers (3 males and 1 female) were excluded from the study because of a degree of hearing loss that fell outside of the range defined for this study. Data were collected and analyzed from the remaining 33 participants (10 males and 23 females) who met the inclusion criteria for hearing, middle ear, and cognitive status. Mean pure-tone audiometric thresholds by frequency and ear for the male and female participants are displayed in Table 2 and Figure 1.

Auditory Processing Tests

Preliminary analyses using one-way analysis of variance (ANOVA) revealed no significant effect of gender on the test score ($p > 0.05$) and no significant effect of ear on the test score ($p > 0.05$). Subsequently, data were collapsed for gender and ear for all remaining analyses. Paired samples t-tests determined there was no significant difference in the test outcome when scoring the CST using the Willeford and Burleigh (1994) method as compared with the Musiek and Pinheiro (1985) method in any test condition; therefore, remaining analyses were completed using scores from the Willeford and Burleigh (1994) method only.

A $3 \times 2 \times 2$ repeated-measures factorial ANOVA (test \times length \times repetitions) was used to determine the effects of test, length, and repetitions on the outcome score for all three tests of AP ability. There was a significant main effect of test [$F_{(2, 3)} = 196.44, p < 0.0001$] on the overall score. Specifically, mean scores were

the highest for the CST (99.51%), followed by the FPT (88.47%). Participants scored the lowest on the LPFST, with a mean score of 45.64%.

Repeated-measures factorial ANOVA additionally revealed a main effect of repetition [$F_{(1, 32)} = 73.83, p < 0.0001$] on the test score (Figure 2), as well as interactions between the test and use of repetitions, [$F_{(2, 31)} = 39.80, p < 0.0001$], and the number of test items and use of repetitions [$F_{(1, 32)} = 8.48, p = 0.006$]. Individual repeated-measures factorial ANOVAs were conducted for each test to better explore these interactions. The sphericity assumption was violated in this original (omnibus), so all subsequently reported significance values are the Greenhouse–Geisser-corrected estimates.

FPT

A 2×2 (length \times repetitions) repeated-measures two-way ANOVA revealed a significant main effect of repetition [$F_{(1, 65)} = 45.57, p < 0.0001$], such that mean outcome scores were significantly greater overall when repetitions were used (92.77%) compared with when repetitions were not used (84.16%). No significant effect of length [$F_{(1, 65)} = 0.00145, p = 0.97$] and no interactions between number of test items and use of repetitions [$F_{(1, 65)} = 0.12999, p = 0.72$], were found.

CST

A 2×2 (length \times repetitions) repeated-measures two-way ANOVA revealed no significant effect of repetition [$F_{(1, 65)} = 2.781, p = 0.10$] on the test score. In addition, there was no main effect of length [$F_{(1, 65)} = 0.278, p = 0.60$] on the CST outcome score. There was no interaction between the number of test items and use of repetitions [$F_{(1, 65)} = 2.779, p = 0.10$].

Table 1. Test-Specific Procedures and Associated Norms

| Test | Standard Procedure | Adapted Procedure | Adult Norms |
|-------|--|---|--------------------------|
| FPT | Thirty test items delivered monaurally at a presentation level of 50-dB SL re: 1000-Hz threshold. No repetitions allowed. | Fifteen test items delivered at the same presentation level as outlined in the standard procedure. Repetitions were allowed. | $\geq 80\%$ for each ear |
| CST | Ten sentences with the right ear as the target ear and 10 sentences with the left ear as the target ear. Target sentence is presented at 35-dB SL re: PTA; competing sentence is presented at 50-dB SL re: PTA. No repetitions allowed. | Five target sentences per ear, delivered at the same presentation level as outlined in the standard procedure. Repetitions were allowed. | $\geq 90\%$ for each ear |
| LPFST | Fifty words delivered monaurally at a presentation level of 50-dB HL. No repetitions allowed. | Twenty-five test items delivered at the same presentation level as outlined in the standard procedure. Repetitions were allowed. | $\geq 78\%$ for each ear |

Note: The standard and adapted procedures were both compared with the adult normative data associated with each test. PTA = pure-tone average.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 2. Mean Pure-Tone Air Conduction Thresholds per Ear for Male and Female Participants

| Gender | Ear | 500 Hz | 1000 Hz | 2000 Hz | 4000 Hz | 8000 Hz |
|--------|-------|--------------|--------------|--------------|--------------|---------------|
| Male | Right | 14 (7.38) | 17 (6.75) | 15.5 (8.32) | 15 (8.82) | 23.5 (14.54) |
| | Left | 14.5 (6.85) | 16 (6.58) | 16.5 (7.84) | 17 (5.87) | 26 (12.65) |
| Female | Right | 12.30 (5.81) | 11.96 (6.35) | 13.26 (6.50) | 12.83 (6.88) | 19.35 (13.25) |
| | Left | 12.39 (6.55) | 10.65 (5.29) | 12.39 (4.74) | 13.91 (7.97) | 17.83 (11.16) |

Note: The mean pure-tone thresholds in dB HL by frequency and ear for male and female participants are displayed in the table shown above, with SD values contained in parentheses.

LPFST

A 2 × 2 (length × repetitions) repeated measures two-way ANOVA revealed a significant main effect of repetitions [$F_{(1, 65)} = 112.40, p < 0.0001$]. There was no significant main effect of length [$F_{(1, 65)} = 0.094, p = 0.78$]; however, there was a significant interaction between the number of test items used and use of repetitions [$F_{(1, 65)} = 18.01, p < 0.0001$]. The interaction is such that the effect of repetitions is greater when using a shorted number of test items versus a standard number of test items, as seen in Figure 3.

DISCUSSION

(C)APD is an impairment in the ability of a person’s central auditory nervous system to appropriately integrate auditory input that affects both children and adults to varying percentages (Stach et al, 1990; ASHA, 2005; O’Beirne et al, 2012). At present, controversy surrounding a diagnosis of (C)APD stems from the lack of a

“gold standard” by which to make the diagnosis, multiple conflicting guidelines for clinical practice, and lack of buy-in to the theoretical construct of auditory processing. Diagnosis is further complicated by inadequate academic and clinical preparation for administering, scoring, and interpreting behavioral tests of (C)APD (Sykes et al, 1997; Chermak et al, 1998; Emanuel, 2002; Bellis, 2006; Chermak et al, 2007). To strengthen the clinical significance of a diagnosis of (C)APD, audiologists must commit to administering these tests in the same manner by which the tests were standardized and from which normative data were developed.

Several studies have emerged in recent years to highlight the need for improved administration and interpretation of tests of (C)APD (Cacace and McFarland, 2013; Turner, 2013; Wilson and Arnott, 2013). Cacace and McFarland (2013) expressed the need for more uniform test procedures related specifically to the assessment of (C)APD. In fact, the sentiment expressed by Cacace and McFarland (2013) is strikingly similar to one articulated by Bellis (2003) a decade prior, when

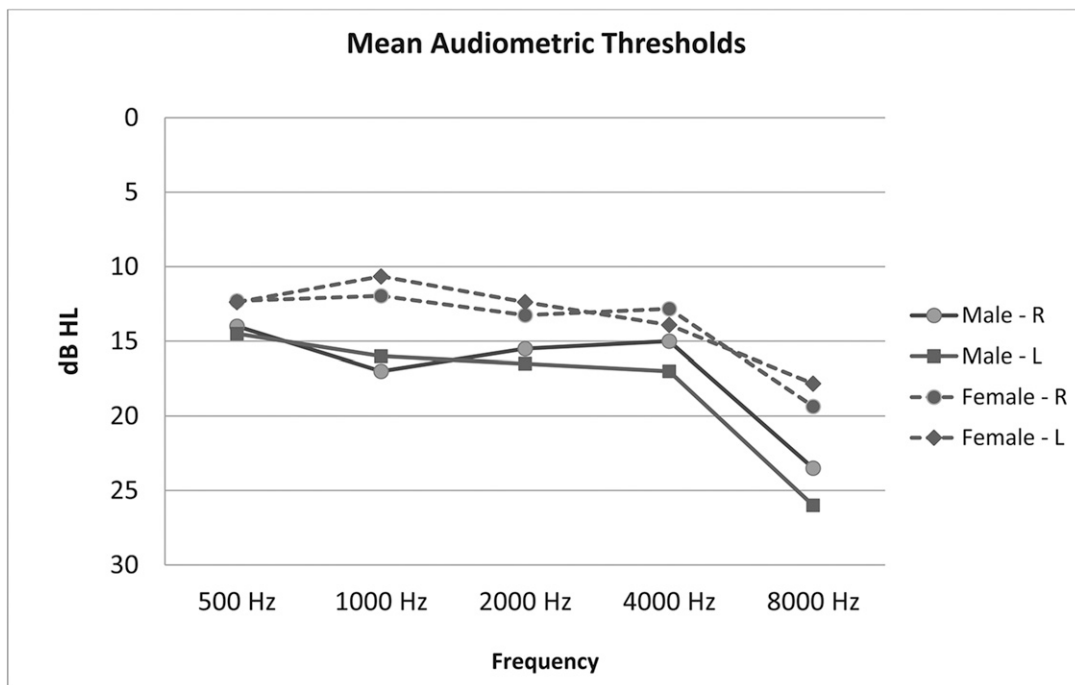


Figure 1. Mean pure-tone thresholds in dB HL by frequency and ear for male and female participants. R = right; L = left.

Mean Outcome Scores With and Without the Use of Repetitions

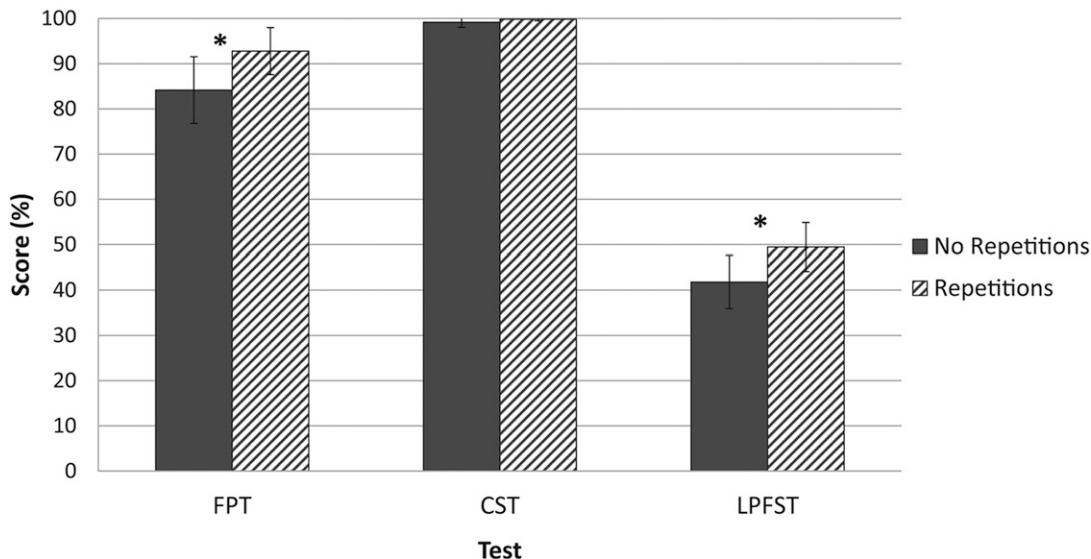


Figure 2. Mean outcome scores for each test are represented across two conditions: without repetitions of incorrectly answered test items and including repetitions of incorrectly answered test items. Error bars represent two standard errors above and below the mean. * represents significant difference between repetition conditions ($p < 0.05$).

she stated that external test factors must be held constant to improve the overall validity of the test. Wilson and Arnott (2013) argued that a diagnosis of (C)APD should be made only when the audiologist explicitly states which set of diagnostic criteria was used to make the diagnosis. They propose that some of the apparent overdiagnosis of (C)APD may be attributable to individual clinicians diagnosing the disorder based on different

diagnostic criteria. The present study was completed in an effort to determine the consequences of modifying standard test procedures and thereby highlight how such practices may contribute to the overall variability and controversy surrounding (C)APD. Subsequently, this study is a means to reemphasize previous calls to maintain test standardization to reduce variability and improve diagnostic validity.

Interaction Between Number of Test Items & Use of Repetitions on LPFST Score

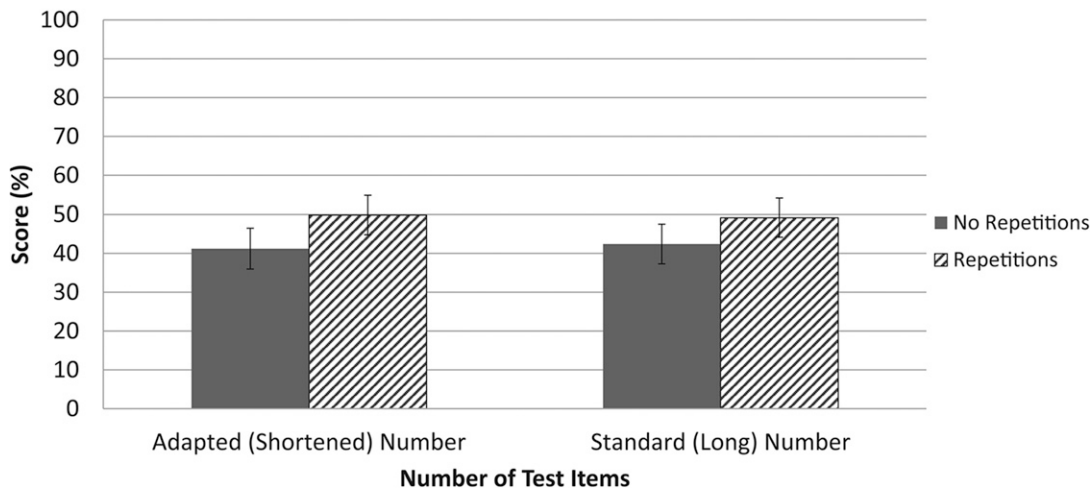


Figure 3. A bar graph depicting the interaction between the number of test items used and use of repetitions. Error bars represent two standard errors above and below the mean. The relationship depicted is such that that there was a greater effect of repetition usage for the adapted (shortened) number of test items compared with the standard (longer) number of test items.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Several clinically relevant findings have emerged from the present study. First, the results of this study support that there is no significant difference in the CST outcome score when calculated using the Willeford and Burleigh (1994) method as opposed to the Musiek and Pinheiro (1985) method. This lack of significance suggests that clinicians can use the scoring method with which they feel most comfortable without sacrificing test validity. It should be noted, however, that the lack of effect found may be due to the perceived ease of the CST for older adults and its resulting ceiling effects. Therefore, this finding should be reexamined in future, larger scale studies.

Second, participants overall scored significantly worse on the filtered speech test. Previous studies have repeatedly shown that older adults have more difficulty understanding speech that has been degraded by the presence of competing background noise (see Atcherson et al, 2015 for a review). Filtered speech stimuli have likewise been degraded, by filtering out important frequency information. Although previous studies have not examined filtered speech performance in older adults, the same ability, auditory closure, underlies both speech-in-noise and filtered speech perception. Poorer peripheral hearing, central presbycusis, and reduced cognitive resources have been proposed as possible explanations for this decline in auditory closure ability with age.

Determining that there was no significant effect of length on any of the three tests of (C)APD used in the present study was surprising, but is promising to both clinicians and patients if confirmed in future studies. A true lack of effect may suggest that shorter versions of individual tests may be effective and should be developed and normed. Alternatively, shorter versions of some auditory processing tests may be useful as screening tests for (C)APD. Either of these alternatives could shorten the length of time needed to administer the entire (C)APD test battery without sacrificing reliability. AAA (2010) has recommended that assessment of (C)APD should last no longer than 45 to 60 minutes in an effort to limit patient fatigue, inattention, and decline in motivation; in shortening the time of the (C)APD test battery, clinicians can strive to be more compliant with these recommendations. A reduction in the number of test items, however, increases the critical difference needed to identify statistically different performance across test sessions and thereby document improvement in ability due to rehabilitation and/or maturation (Thornton and Raffin, 1978).

Unlike varying test length, repeating incorrectly answered test items did show a significant effect on the test outcome. For two of the three tests in the present study, allowing the patient to respond a second time to test items initially answered incorrectly yielded a statistically significant change in the participant's overall

adjusted test score compared with his or her original score. This finding is supported by research that has shown that adults are more likely to guess unknown answers than children, making responses from adults more variable than responses from children (Rowley and Traub, 1977; Huff et al, 2011). By allowing adults the chance to repeat their answers and thereby increasing the variability of their responses, the use of repetitions in the present study was found to significantly alter test outcome for the FPT and LPFST. The authors of the present study would propose that there are several scenarios that could occur when repetitions of incorrectly answered test items are included in testing:

- The person did not hear the stimulus correctly during its first presentation and answered incorrectly. On repetition of the stimulus, the person will answer correctly.
- The person truly has (C)APD and believed his/her incorrect answer to be correct during the first presentation of the stimulus. The individual will again answer incorrectly when the stimulus is presented for the second time.
- The person guessed the answer incorrectly during the first presentation. When the stimulus is presented for a second time, the individual will guess again. He or she will either guess the answer correctly because of chance (Seashore et al, 1954), or the person will again guess incorrectly.

Based on the findings in the present study that scores improved when the test score was adjusted for repetition use, it is the assumption of the authors that one or both of the following occurred:

- The person did not hear the stimulus correctly during its first presentation and answered incorrectly. On repetition of the stimulus, the person answered the item correctly.
- The person guessed the answer incorrectly during the first presentation. When the stimulus was presented for a second time, the person guessed the answer correctly because of chance (Seashore et al, 1954).

It is unclear how often each of these two scenarios occurred in the present study, and thereby, the authors are unable to make a statement regarding whether or not the use of repetitions of incorrectly answered test items lends itself to a more accurate diagnosis of (C)APD. What is clear, however, is the following implication: if a clinician decides to repeat incorrectly answered test items and if this is not explicitly allowed according to test standardization materials, there is a possibility that the test outcome may be falsely elevated compared with the participant's true score. Although to

the authors' knowledge no test permits the use of repetitions, tests of (C)APD often do not come with literature explicitly detailing the procedures by which to administer the tests (Emanuel, 2002). This may be problematic when these tests are administered by audiologists with limited experience specific to the assessment of (C)APD, who may be more likely to unintentionally vary from standard procedures. Furthermore, it is not unlikely to assume that variances in the preferred test administration methods will nonetheless occur, regardless of the tester's experience, as these variances have been documented to occur in relation to other tests frequently completed as part of a basic audiological evaluation (Schoepflin, 2012; Hornsby and Mueller, 2013).

Finally, there was a significant interaction between the length of test and repetition for the LPFST. This illustrates the interrelated nature of various test modifications and the complexity associated with test administration. Changing the length or repetitions allowed for each test may seem like a simple modification, but may affect test outcome in unforeseen and/or unpredictable ways that are not evident in this study. Reducing test length, as an example, can change the phonetic balance of word lists for a test like LPFST and may, therefore, reduce the validity of the test. Further study is needed before any changes to present clinical tests can be implemented.

Overall, this study determined that varying certain test administration procedures alone (e.g., repeating missed test items) and in conjunction with other variances (e.g., repeating missed test items and varying the length of the test) can affect the test outcome. Because a diagnosis of (C)APD is made after an individual scores below normal limits on only two or more tests of (C)APD, changing the outcome on even one test from a "pass" to a "fail," or vice versa, could ultimately affect whether or not a diagnosis of (C)APD is made.

LIMITATIONS AND FUTURE RESEARCH

Because of time constraints, this study was limited to the evaluation of the auditory processing abilities of 33 older adults varying in age from 50 to 65 years, using only three tests of auditory processing ability. In the future, this study should be repeated with a larger sample of adults with a greater age range. As all adults in this study were recruited through Towson University, a more diverse group of participants should be used in the future. In addition, different tests of auditory processing ability should be used in subsequent studies. Although the prevalence of (C)APD is higher in older adults (Stach et al, 1990; Golding et al, 2004), auditory processing is most frequently evaluated in school-age children. As such, it is recommended that a similar study be conducted on the pediatric population

to determine if the results found in the present study hold true when assessing children. One notable change in results when assessing children would likely be the presence of an ear effect on test scores. The lack of ear effect in the present study is unsurprising, as a right-ear advantage is often only seen in young children before the full maturation of the corpus callosum (Bellis, 2003; Whitelaw, 2008). This study evaluated the auditory processing abilities of older adults in whom the corpus callosum is expected to be fully myelinated, and therefore, no ear effect was expected. Should this study be completed on young children, results may yield a different outcome.

SUMMARY

This study confirms that varying test administration procedures has the ability to affect the outcome of one or more tests of (C)APD and, therefore, the potential to change the resulting diagnosis of (C)APD. Unclear test procedures, in addition to the lack of a "gold standard" diagnostic measure and the limited widespread use of objective assessment methods, fuel the controversy currently related to (C)APD. This is further complicated when the potential lack of adequate clinical and academic exposure to (C)APD is considered. To reduce the variability and increase the clinical significance of (C)APD as a true, standalone disorder, audiologists must adhere to the test administration and scoring procedures originally used to standardize the test.

REFERENCES

- American Academy of Audiology (AAA). (2010) Clinical Practice Guidelines: Diagnosis, Treatment and Management of Children and Adults with Central Auditory Processing Disorder. <http://www.audiology.org/resources/>. Accessed June 12, 2012.
- American Speech-Language-Hearing Association (ASHA). (2005) (Central) auditory processing disorders [Technical Report]. www.asha.org/policy. Accessed June 12, 2012.
- Atcherson SR, Nagaraj NK, Kennett SEW, Levisse M. (2015) Overview of central auditory processing deficits in older adults. *Semin Hear* 36(3):150–161.
- Bellis TJ. (2003) *Assessment and Management of Central Auditory Processing Disorders in the Educational Setting: From Science to Practice*. 2nd ed. Clifton Park, NY: Delmar Cengage Learning.
- Bellis TJ. (2006) Audiologic behavioral assessment of (C)APD. In: Parthasarathy TK, ed. *An Introduction to Auditory Processing Disorders in Children*. New York, NY: Lawrence Erlbaum Associates, 63–80.
- British Society of Audiology (BSA). (2011) *Position Statement on APD*. <http://www.thebsa.org.uk/>. Accessed May 17, 2017.
- Cacace AT, McFarland DJ. (2013) Factors influencing tests of auditory processing: a perspective on current issues and relevant concerns. *J Am Acad Audiol* 24(7):572–589.

- Chermak GD, Silva ME, Nye J, Hasbrouck J, Musiek FE. (2007) An update on professional education and clinical practices in central auditory processing. *J Am Acad Audiol* 18:428–452.
- Chermak GD, Traynham WA, Seikel JA, Musiek FE. (1998) Professional education and assessment practices in central auditory processing. *J Am Acad Audiol* 9:452–465.
- Chermak GD. (2016) 20Q: (C)APD—Fundamentals. *AudiologyOnline*. Article #17765. <http://www.audiologyonline.com/>. Accessed May 17, 2017.
- Cox LC, McCoy SL, Tun PA, Wingfield A. (2008) Monotic auditory processing disorder tests in the older adult population. *J Am Acad Audiol* 19(4):293–308.
- Dawes P, Bishop D. (2009) Auditory processing disorder in relation to developmental disorders of language, communication, and attention: a review and critique. *Int J Lang Commun Disord* 44(4):440–465.
- Emanuel DC, Ficca KN, Korczak P. (2011) Survey of the diagnosis and management of auditory processing disorder. *Am J Audiol* 20(1):48–60.
- Emanuel DC. (2002) The auditory processing battery: survey of common practices. *J Am Acad Audiol* 13:93–117.
- Folstein MF, Folsetin SE, McHugh PR. (1975) “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12(3):189–198.
- Friberg JC, McNamara TL. (2010) Evaluating the reliability and validity of (central) auditory processing tests: a preliminary investigation. *J Educ Audiol* 16:4–17.
- Golding M, Carter N, Mitchell P, Hood L. (2004) Prevalence of central auditory processing (CAP) abnormality in an older adult Australian population: the Blue Mountains hearing study. *J Am Acad Audiol* 15:633–642.
- Hornsby B, Mueller HG. (2013) Monosyllabic word testing: five simple steps to improve accuracy and efficiency. *AudiologyOnline*, Article #11978. <http://www.audiologyonline.com/>. Accessed May 17, 2017.
- Huff MJ, Meade ML, Hutchison KA. (2011) Age-related differences in guessing on free and forced recall tests. *Memory* 19(4):317–330.
- Katz J, Johnson CD, Tillery KL, Bradham T, Brandner S, Delagrangé TN, Ferre JM, King J, Kossover-Wechter D, Lucker JR, Medwetsky L, Saul RS, Rosenberg GG, Stecker NA. (2002) Clinical and research concerns regarding Jerger & Musiek (2000) APD recommendations. *AudiologyOnline*, Article #1182. <http://www.audiologyonline.com/>. Accessed June 12, 2012.
- Kelly A. (2007) Normative data for behavioural tests of auditory processing for New Zealand school children aged 7 to 12 years. *Aust N Z J Audiol* 29:60–64.
- McDermott EE, Smart JL, Boiano JA, Bragg LE, Colon TN, Hanson EM, Emanuel DC, Kelly AS. (2016) Assessing auditory processing abilities in typically developing school-aged children. *J Am Acad Audiol* 27(2):72–84.
- Musiek FE. (1994) Frequency (pitch) and duration pattern tests. *J Am Acad Audiol* 5(4):265–268.
- Musiek FE. (2002) The frequency pattern test: a guide. *Hear J* 55(6):58.
- Musiek FE, Chermak GD, Weihing J, Zappulla M, Nagle S. (2011) Diagnostic accuracy of established central auditory processing test batteries in patients with documented brain lesions. *J Am Acad Audiol* 22(6):342–358.
- Musiek FE, Pinheiro ML. (1985) Dichotic speech tests in the detection of central auditory dysfunction. In: Pinheiro M, Musiek F, eds. *Assessment of Central Auditory Dysfunction: Foundations and Clinical Correlates*. Baltimore, MD: Williams & Wilkins, 201–218.
- Musiek FE, Pinheiro ML. (1987) Frequency patterns in cochlear, brainstem, and cerebral lesions. *Audiology* 26:79–88.
- O’Beirne GA, McGaffin AJ, Rickard NA. (2012) Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *Int J Pediatr Otorhinolaryngol* 76:777–782.
- Rowley GL, Traub RE. (1977) Formula scoring, number-right scoring, and test-taking strategy. *J Educ Meas* 14(1):15–22.
- Schoepflin JR. (2012) Back to basics: speech audiometry. *AudiologyOnline*, Article #6828. <http://www.audiologyonline.com/>. Accessed May 17, 2017.
- Seashore HG, Wesman AG, Doppelt JE, Gelink M, Ricks Jr, JH. (1954) *The Correction for Guessing*. Test Service Bulletin No. New York, NY: The Psychological Corporation.
- Stach BA, Spretnjak ML, Jerger J. (1990) The prevalence of central presbycusis in a clinical population. *J Am Acad Audiol* 1(2):109–115.
- Stewart BS. (2003) The word intelligibility by picture identification test: a two-part study of familiarity and use. *J Educ Audiol* 11:39–48.
- Sykes S, Tucker D, Herr D. (1997) Aural rehabilitation and graduate audiology programs. *J Am Acad Audiol* 8(5):314–321.
- Thornton A, Raffin M. (1978) Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res* 21(3):507–518.
- Tomlin D, Dillon H, Sharma M, Rance G. (2015) The impact of auditory processing and cognitive abilities in children. *Ear Hear* 36(5):527–542.
- Turner RG. (2013) Understanding protocol performance: impact of test performance. *J Am Acad Audiol* 24(10):909–919.
- Whitelaw GM. (2008) Assessment and management of auditory processing disorders in children. In: Madell JR, Flexer C, eds. *Pediatric Audiology: Diagnosis, Technology, and Management*. New York, NY: Thieme Medical Publishers, Inc., 145–155.
- Wiley TL, Stoppenbach DT, Feldhake LJ, Moss KA, Thordardottir ET. (1995) Audiologic practices: what is popular versus what is supported by evidence. *Am J Audiol* 4(1):26–34.
- Willeford JA, Burleigh JM. (1994) Sentence procedures in central testing. In: Katz J, ed. *Handbook of Clinical Audiology*. 4th ed. Baltimore, MD: Williams & Wilkins, 256–268.
- Wilson WJ, Arnott W. (2013) Using different criteria to diagnose (central) auditory processing disorder: how big a difference does it make? *J Speech Lang Hear Res* 56:63–70.