## ACTIVITY: Test Validity Rests in Evidence, Part 2 Answer Guide

### Instructions:

The **goal** of this activity is to give you experience with an in-depth analysis and comparison of two diagnostic accuracy studies. You will complete it in two parts. In Part 1, you will identify two diagnostic accuracy studies and complete a comparison table. In Part 2, you will compare your findings. Follow the steps for each part below.

Example answers for Part 2 are below. Student answers will depend on topic and articles selected.

### Relevant articles for this assignment:

- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American*, 283(4), 82-87.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., ... & STARD Group\*. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Annals of internal medicine*, *138*(1), 40-44.
- Vermiglio, A. J. (2014). On the clinical entity in audiology: (Central) auditory processing and speech recognition in noise disorders. *Journal of the American Academy of Audiology*, *25*(09), 904-917.
- Vermiglio, A. J. (2016). On diagnostic accuracy in audiology: Central site of lesion and central auditory processing disorder studies. *Journal of the American Academy of Audiology*, *27*(02), 141-156. (With errata)
- Vermiglio, A. J., & Fang, X. (2022). The World Health Organization (WHO) hearing impairment guidelines and a speech recognition in noise (SRN) disorder. *International Journal of Audiology*, *61*(10), 818-825. [Note: This article is basically a tutorial on diagnostic accuracy studies.]
  [Note: Do not confuse reliability with validity. Reliability refers to the repeatability of a test. <u>Validity</u> refers to the ability of a test to measure what it is supposed to measure.]

#### Part 1

 Find a diagnostic accuracy study on a test used in audiology or speech-language pathology for the detection of a specific target disorder. Recall from the video lecture that the terms diagnostic accuracy, index test, target disorder, and "gold" or reference standard test may not appear in these types of articles.

Suggested search strategies include using Boolean cues in Google Scholar.

For example: "sensitivity" AND "specificity" AND "name of index test"

"sensitivity" AND "specificity" AND "name of the target disorder"

At minimum the article should report the sensitivity and specificity of the index test or tests, it should contain a control group and at least one disordered group and it should report the index

test results including the means, standard deviations, and the number of subjects.

- 2. Find a second article with the same index test and/or target disorder as the first article for comparison.
- 3. Use the information from the article(s) to complete the data fields in the Excel Table template. *Data Field Descriptions*

Author(s) (year) – article citation

Reference – complete article reference

Diagnostic system - index test, target disorder, "gold" or reference standard test

**Index test** – the test under evaluation. This may appear in the form of an index test battery with multiple tests.

**Target disorder** – This may also be called a clinical entity (see Vermiglio, 2014) or diagnostic target.

**Index test cut-point** – This value delineates between the positive (disorder is presence) and negative (disorder is absent) index test results.

**"Gold" or reference test** – This is the considered the best method for the identification of the presence and absence of the target disorder. In a diagnostic accuracy study, this is the test (or test battery) used to sort research participants into the control and disordered groups.

**"Gold" or reference test cut-point** – This value delineates between participants with and without the target disorder.

**Control group** – The group without the target disorder as identified by the "gold" or reference standard test.

**Disordered group** – The group with the target disorder as identified by the "gold" or reference standard test.

**Index test group difference** – This is the most elemental calculation in a diagnostic accuracy study. A statistically significant index test result group difference indicates that, to some degree, the index test can delineate between participants with and without the disorder.

Sensitivity – This is the same as the percentage of true positive index test results.

Specificity – This is the same as the percentage of true negative index test results.

Convergent validity – This describes relationship between index test vs "gold" or reference

standard test results. A strong and statistically significant relationship indicates that the results of the reference standard test may be predicted from the index test results.

**Area under the curve (AUC)** – This describes the strength of the index test as a predictor of the target disorder. An AUC of 1 indicates perfect separation of the participants into groups with and without the disorder. An AUC of 0.5 indicates that the index test has a 50/50 chance of making a correct diagnosis of the target disorder (no better than flipping a coin). See Swets et al. (2000) for an explanation.

## <u>Part 2</u>

Write a report that compares and analyzes the findings from the diagnostic studies.

### 1. Introduction

- a. Introduce the topic and explain why it is relevant for audiologists and/or speech-language pathologists. Include the appropriate article citations.
   ANSWER: Audiologists and speech-language pathologists diagnose and treat disorders. It is imperative that the results of the diagnostic tests are accurate to provide correct diagnoses and for the selection of the appropriate intervention or treatment.
- b. Give some background information on the purpose of diagnostic accuracy studies. ANSWER: Diagnostic accuracy studies are used to determine the ability of an index test to correctly identify the presence and absence of a target disorder (Bossuyt et al., 2003). Results from diagnostic accuracy studies may be used to inform audiologists and/or speech-language pathologists about the suitability of an index test for use in the clinic and in research.
- c. Describe the components of a diagnostic system.
   ANSWER: The diagnostic system includes the index test (or the test under evaluation), the target disorder, and the "gold" or reference standard test.
- d. Describe how the diagnostic system is used to determine test validity. ANSWER: In a diagnostic accuracy study, the "gold" or reference standard test is used to divide the study participants into groups with and without the disorder. The index test results are classified as *positive* or *negative* for the presence of the target disorder. The index test results are compared to the results of the "gold" or reference standard test to determine if the index test results are *true* or *false*.
- Explain why it is important for a clinical audiologist or speech-language pathologist to know the diagnostic accuracy or validity of their diagnostic tests.
   ANSWER: The utilization of diagnostic tests with poor or unknown diagnostic accuracy may lead to misdiagnoses and unnecessary treatment.

## 2. Study Overview: Article 1

Describe the goal of first research article. Give some information on the target disorder. Explain the importance of identifying the presence of this disorder. Include the relevant citations.

 $\rightarrow$  EXAMPLE ARTICLE 1: Diagnostic Accuracy of the Hering in Noise Test for the Detection of a Speech Recognition in Noise Disorder

ANSWER: While pure-tone thresholds have been used as a "gold standard" measure for the assessment of speech recognition in noise ability (WHO, 2021), it has been shown to be relatively insensitive to a speech recognition in noise disorder (Middelweerd et al., 1990). This means that a speech recognition in noise disorder must be measured directly and not inferred from pure-tone thresholds. It is preferable to measure speech perception in noise ability with a diagnostically accurate speech recognition in noise test. Vermiglio et al. (2018) investigated the diagnostic accuracy of the Hearing in Noise Test or HINT (Nilsson et al., 1994; Vermiglio, 2008) for the detection of a speech recognition in noise disorder.

# 3. Components of the Diagnostic System(s): Article 1

- a. Index Test(s) or Index Test Battery:
  - i. Give the details of the index test or test battery.

ANSWER: The HINT was used as the index test. HINT sentences were presented in the presence of steady-state noise. The noise was presented at a fixed level of 65 dBA and the sentences were adaptively varied in level based on the participant's response. The stimuli were delivered via headphones in a virtual sound environment. The step-size for the first 4 sentences was 4 dB and 2 dB for the remaining 20 sentences. The thresholds were expressed as the dB signal-tonoise ratio (SNR) for the correct identification of 50% of the sentences.

- Was the index test or index test battery described clearly enough to allow for replication of the study?
   ANSWER: Yes, the information on the test parameters and procedures allows for the replication of this study.
- iii. Include a table showing the diagnostic system(s) used in the study. ANSWER in table below.

# Table 1

Diagnostic Systems Used in Article 1

Author, year	Diagnostic System	Index Test	Target Disorder	"Gold" or Reference Standard Test
Vermiglio et al. (2018)	1	HINT (average threshold across 3 conditions; noise front, noise right, noise left)	Speech recognition in noise disorder	Self-report
	2			
	3			
	4			

b. Target Disorder:

Is the target disorder a clinical entity according to the Sydenham-Guttentag criteria? Briefly describe each criterion and how it relates to the target disorder (see Vermiglio, 2014 for two examples).

ANSWER: Yes, a speech recognition in noise disorder is a clinical entity or legitimate disorder according to the Sydenham-Guttentag criteria. It has an unambiguous definition. It represents a homogenous patient group. It represents a limitation for the patient and it facilitates diagnosis and intervention (Vermiglio, 2014).

- c. "Gold" or Reference Standard Test:
  - i. Did the author(s) give any evidence showing that the "gold" or reference standard test is valid? In other words, did they indicate that the "gold" or reference standard test is the best way or one of the best ways to identify the presence and absence of the target disorder? For example, was the reference standard test used as an index test in a prior study showing good diagnostic accuracy (sensitivity, specificity) of this test?
    ANSWER: The "gold" or reference standard test in this study was the self-report of speech perception in noise difficulties. The authors argued that self-report has been used as a "gold" or reference standard test for such disorders as pain (Stilma et al., 2015), tinnitus (Schaette & McAlpine, 2011), hearing loss (Beasley,

1940), and a speech recognition in noise disorder (Middelweerd et al., 1990).

ii. Was the "gold" or reference standard test described clearly enough to allow for replication?ANSWER: Yes, the "gold" standard test was simply the self-report of the

participant's ability to recognize speech in a noisy environment such as a crowded restaurant. This allows for replication.

- d. Participant Characteristics:
  - What was the inclusion criteria for the study participants? ANSWER: All participants were native speakers of American English, and all had normal pure-tone thresholds (< 25 dB HL, 250 – 6000 Hz).</li>
  - ii. How many participants were included in each group? ANSWER: There were 22 participants in the control group and 25 participants in the disordered group.
  - iii. What was the mean age of the participants? ANSWER: The mean age for the control group was 36.91 years. The mean age for the participants in the disordered group was 36.24 years.
  - iv. Was the index test and reference standard test administered to all subjects in the control (non-disordered) and disordered group?
     ANSWER: Yes.
  - v. Did the control group include those without the disorder? ANSWER: Yes, all participants in the control group reported no difficulty with

speech perception in noisy environments.

- vi. Did the disordered group contain those with the disorder? ANSWER: Yes, all participants in the disordered group reported difficulty with speech perception in noisy environments
- vii. Did the control or disordered subjects have other disorders or conditions that may have affected the index and reference standard test results?
   ANSWER: No disorders or conditions were reported by the participants.

### 4. Results: Article 1

- a. Enter the appropriate results in Tables 2 and 3 below. SEE ANSWERS IN TABLES BELOW
- b. Describe the descriptive statistics (from Table 2) for the index test and reference standard test results (mean and SD) for each group (required). ANSWER: The control group performed better than the disordered group for the HINT average threshold. This 1.15 dB difference was statistically significant (p < 0.01). A wider range of thresholds was found for the disordered group than the control group.
- c. Describe the minimum maximum, range, (if available). Note: if not provided, this may be estimated from a scatter plot.
   ANSWER: The minimum average HINT threshold for the control group was -9.60 dB SNR. This is better than the control group's maximum threshold of -6.87 dB SNR. The range of average HINT thresholds for the control group was 2.73 dB. The minimum average HINT threshold was for the disordered group was -8.73 dB SNR. This is a better than the maximum threshold of -4.67 dB SNR. The range of average HINT thresholds for the control group was -8.73 dB SNR. This is a better than the maximum threshold of -4.67 dB SNR. The range of average HINT thresholds for the control group was 4.06 dB.
- d. Describe the sensitivity and specificity of the index tests (from Table 3). ANSWER: The sensitivity of the index test for the detection of a speech recognition in noise disorder is 80%. This means that the index test correctly identified 80% of the research participants with the target disorder. The specificity of the index test for the detection of the absence of a speech recognition in noise disorder is 86%. This means that the index test correctly identified 86% of the research participants without the target disorder.
- e. Describe the group differences in index test performances and *p*-values if available. ANSWER: The group difference in index test performances was 1.15 dB. The control group performed better than the disordered group. This difference was statistically significant (p < 0.01).
- f. Describe the correlation coefficients between the reference standard test vs. the index test results and the *p*-values (if available).
   ANSWER: The result of reference standard test is not a continuous variable. Therefore it

# is not possible to calculate this statistic.

- g. Do the correlation coefficients (if available) indicate that the reference standard test results can be predicted from the index test result?
   ANSWER: Not applicable.
- h. Describe the receiver operating characteristics (ROC) curve(s) and the area under the curve (AUC; if available).
   ANSWER: A figure of the ROC curve was not shown in this study for the index test (average HINT threshold). However, the AUC for this index test was 0.86 (p < 0.01). This indicates that the index test is a statistically significant predictor of the target disorder.</li>

# Table 2

# Article 1 Descriptive Statistics

Author, year	Group	Test	Mean (unit of measure)	Standard Deviation	Maximum Result (unit of measure)	Minimum Results (unit of measure)	Range (e.g., dB or percentage points)
Vermiglio	Control	Index HINT	-8.30 dB SNR	0.66	-6.87 dB SNR	-9.60 dB SNR	2.73 dB
(2018)	Disordered	(average threshold)	-7.16 dB SRN	0.98	-4.67 dB SNR	-8.73 dB SNR	4.06 dB
	Control	Gold	Not				
	Disordered	Standard	applicable				

# Table 3

## Article 1 Results of the Diagnostic Accuracy Study

Author, year	Diagnostic System	Index Test	Sensitivity	Specificity	Index Test vs. Reference Standard Test Results, Correlation Coefficients (p-values in parentheses)	Area Under the Curve (p- value)
Vermiglio et al. (2018)	1	HINT (average threshold)	80%	86%	Not applicable	AUC = .86 (p < 0.01)
	2					
	3					
	4					

# 5. Discussion: Article 1

Discuss the implications of the results for your work as a clinician and/or researcher. Include the appropriate citations throughout.

ANSWER: The results of this study indicate that the index test under evaluation would be a reasonable addition to a clinical protocol. These results contradict the hearing guidelines from

the World Health Organization which state that speech perception in noise may be predicted from pure-tone threshold average (WHO, 2021). According to the WHO, none of the participants in the present study should have difficulties with speech perception in noise. However, the results of this study show a measurable.

### 6. Critique: Article 1

- Describe any limitations or shortcomings of the article.
   ANSWER: This study did not include participants with elevated pure-tone thresholds.
   Future investigations should evaluate the effect of hearing loss on the diagnostic accuracy of the index test.
- Describe anything that was unclear in the article.
   ANSWER: Details of the virtual sound field presentation under headphones was not described.
- c. Describe the generalizability of the study results. Are there limitations in using the information from this study to your clinical patients?
   ANSWER: These results may be generalized for Native English speakers in the same age range as the study participants. They may not be relevant for younger and older participants and those who are non-native English speakers.

## 7. Study Overview: Article 2

Describe the goal of second research study. Give some information on the target disorder. Explain the importance of identifying the presence of this disorder. Include the relevant citations.

 $\rightarrow$  Example Article 2: Diagnostic Accuracy of Pure-Tone Threshold Average for the Detection of a Speech Recognition in Noise Disorder

ANSWER: While pure-tone thresholds have been used as a "gold standard" measure for the assessment of speech recognition in noise ability (WHO, 2021), it has been shown to be relatively insensitive to a speech recognition in noise disorder (Middelweerd et al., 1990). This means that a speech recognition in noise disorder must be measured directly and not inferred from pure-tone thresholds. It is preferable to measure speech perception in noise ability with a diagnostically accurate speech recognition in noise test. Vermiglio et al. (2018) investigated the diagnostic accuracy of the Hearing in Noise Test or HINT (Nilsson et al., 1994; Vermiglio, 2008) for the detection of a speech recognition in noise disorder.

## 8. Components of the Diagnostic System(s): Article 2

- a. Index Test(s) or Index Test Battery:
  - i. Give the details of the index test or test battery.
    - ANSWER: The bilateral pure-tone threshold average (BPTA) was used as the index test. Pure-tone thresholds were obtained for 250 6000 Hz. The BPTA was determined for 500, 1000, and 2000 Hz. These frequencies are commonly

used for PTA in clinical audiology.

- Was the index test or index test battery described clearly enough to allow for replication of the study?
   ANSWER: Yes, the information on the test parameters and procedures allows for the replication of this study.
- iii. Include a table showing the diagnostic system(s) used in the study. ANSWER in table below.

## Table 4

Diagnostic Systems Used in Article 2

Author, year	Diagnostic	Index Test	Target	"Gold" or Reference
	System		Disorder	Standard Test
Vermiglio et	1	Pure-tone threshold testing	Speech	Self-report
al. (2018)			recognition in	
			noise disorder	
	2			
	3			
	4			

b. Target Disorder:

Is the target disorder a clinical entity according to the Sydenham-Guttentag criteria? Briefly describe each criterion and how it relates to the target disorder (see Vermiglio, 2014 for two examples).

ANSWER: Yes, a speech recognition in noise disorder is a clinical entity or legitimate disorder according to the Sydenham-Guttentag criteria. It has an unambiguous definition. It represents a homogenous patient group. It represents a limitation for the patient and it facilitates diagnosis and intervention (Vermiglio, 2014).

- a. "Gold" or Reference Standard Test:
  - i. Did the author(s) give any evidence showing that the "gold" or reference standard test is valid? In other words, did they indicate that the "gold" or reference standard test is the best way or one of the best ways to identify the presence and absence of the target disorder? For example, was the reference standard test used as an index test in a prior study showing good diagnostic accuracy (sensitivity, specificity) of this test?

ANSWER: The "gold" or reference standard test in this study was the self-report of speech perception in noise difficulties. The authors argued that self-report has been used as a "gold" or reference standard test for such disorders as pain (Stilma et al., 2015), tinnitus (Schaette & McAlpine, 2011), hearing loss (Beasley, 1940), and a speech recognition in noise disorder (Middelweerd et al., 1990).

- Was the "gold" or reference standard test described clearly enough to allow for replication?
   ANSWER: Yes, the "gold" standard test was simply the self-report of the participant's ability to recognize speech in a noisy environment such as a crowded restaurant. This allows for replication.
- b. Participant Characteristics:
  - What was the inclusion criteria for the study participants? ANSWER: All participants were native speakers of American English, and all had normal pure-tone thresholds (≤ 25 dB HL, 250 – 6000 Hz).
  - iv. How many participants were included in each group? ANSWER: There were 22 participants in the control group and 25 participants in the disordered group.
  - what was the mean age of the participants?
     ANSWER: The mean age for the control group was 36.91 years. The mean age for the participants in the disordered group was 36.24 years.
  - vi. Was the index test and reference standard test administered to all subjects in the control (non-disordered) and disordered group?
     ANSWER: Yes.
  - vii. Did the control group include those without the disorder?
     ANSWER: Yes, all participants in the control group reported no difficulty with speech perception in noisy environments.
  - viii. Did the disordered group contain those with the disorder? ANSWER: Yes, all participants in the disordered group reported difficulty with speech perception in noisy environments.
  - ix. Did the control or disordered subjects have other disorders or conditions that may have affected the index and reference standard test results?
     ANSWER: No disorders or conditions were reported by the participants.

#### 9. Results: Article 2

- a. Enter the appropriate results in Tables 5 and 6 below. SEE ANSWERS IN TABLES BELOW.
- b. Describe the descriptive statistics (from Table 5) for the index test and reference standard test results (mean and SD) for each group (required).
  ANSWER: The control group performed poorer than the disordered group for the bilateral pure-tone average. This 0.09 dB difference was not statistically significant (*p* = 0.53). A slightly wider range of thresholds was found for the disordered group than the control group.

- c. Describe the minimum maximum, range, (if available). Note: if not provided, this may be estimated from a scatter plot.
  ANSWER: The minimum bilateral pure-tone threshold average for the control group was -0.83 dB HL. This is better than the control group's maximum threshold of 15 dB HL. The range of bilateral pure-tone threshold averages for the control group was 15.83 dB. The minimum bilateral pure-tone threshold average for the disordered group was -1.67 dB HL. This is a better than the maximum of 15 dB HL. The range of average HINT thresholds for the control group was 16.67 dB.
- d. Describe the sensitivity and specificity of the index tests (from Table 6). ANSWER: The sensitivity of the index test for the detection of a speech recognition in noise disorder is 28%. This means that the index test correctly identified 28% of the research participants with the target disorder. The specificity of the index test for the detection of the absence of a speech recognition in noise disorder is 95%. This means that the index test correctly identified 28% of the detection of the absence of a speech recognition in noise disorder is 95%. This means that the index test correctly identified 95% of the research participants without the target disorder.
- e. Describe the group differences in index test performances and *p*-values if available. ANSWER: The group difference in index test performances was -0.09 dB. The control group performed poorer than the disordered group. This difference was not statistically significant (p = 0.53).
- f. Describe the correlation coefficients between the reference standard test vs. the index test results and the *p*-values (if available).
  ANSWER: The result of reference standard test is not a continuous variable. Therefore, it is not possible to calculate this statistic.
- g. Do the correlation coefficients (if available) indicate that the reference standard test results can be predicted from the index test result?
   ANSWER: Not applicable.
- h. Describe the receiver operating characteristics (ROC) curve(s) and the area under the curve (AUC; if available).
  ANSWER: A figure of the ROC curve was not shown in this study for the index test (average HINT threshold). However, the AUC for this index test was 0.51 (p = 0.94). This indicates that the index test is a poo predictor of the target disorder (no better than chance).

## Table 5

Author, vear	Group	Test	Mean (unit of	Standard Deviation	Maximum Result	Minimum Results	Range (e.g., dB or
			measure)		(unit of measure)	(unit of measure)	percentage points)

Vermiglio et al. (2018)	Control	Index Bilateral	4.62 dB HL	3.82	15 dB HL	-0.83 dB HL	15.83 dB
	Disordered	Pure- Tone Threshold Average	4.53 dB HL	4.6	15 dB HL	-1.67 dB HL	16.67 dB
	Control	Gold	Not				
	Disordered	Standard	applicable				

### Table 6

Article 2 Results of the Diagnostic Accuracy Study

Author, year	Diagnostic System	Index Test	Sensitivity	Specificity	Index Test vs. Reference Standard Test Results, Correlation Coefficients (p-values in parentheses)	Area Under the Curve (p- value)
Vermiglio et al. (2018)	1	Pure-tone threshold test (bilateral PTA)	28%	95%	Not applicable	AUC = .51 (p = 0.94)
	2					
	3					
	4					

## 10. Discussion: Article 2

Discuss the implications of the results for your work as a clinician and/or researcher. Include the appropriate citations throughout.

ANSWER: The results of this study indicate that the bilateral pure-tone threshold average is a very poor predictor of a speech recognition in noise disorder. These results contradict the hearing guidelines from the World Health Organization which state that speech perception in noise may be predicted from pure-tone threshold average (WHO, 2021). A speech recognition in noise disorder should be assessed using a diagnostically accurate measure and not inferred from a pure-tone threshold average.

## 11. Critique: Article 2

- a. Describe any limitations or shortcomings of the article.
   ANSWER: This study did not include participants with elevated pure-tone thresholds.
   Future investigations should evaluate the effect of hearing loss on the diagnostic accuracy of the index test.
- Describe anything that was unclear in the article.
   ANSWER: Details of the virtual sound field presentation under headphones was not described.

c. Describe the generalizability of the study results. Are there limitations in using the information from this study to your clinical patients?
 ANSWER: These results may be generalized for Native English speakers in the same age range as the study participants. They may not be relevant for younger and older participants and those who are non-native English speakers.

## 12. Comparison of Two Diagnostic Accuracy Studies

Complete Table 7 to compare the results of each diagnostic accuracy study. ANSWERS IN TABLE BELOW.

## Table 7

Comparison of the Diagnostic Accuracy Results Found in Two Different Diagnostic Accuracy Studies

Author, year	Diagnost ic System	Index Test	Target Disorder	Referenc e Standard Test	Sensitivit Y	Specificit Y	Index Test. Vs. Reference Standard Correlatio n Coefficien ts (p- values)	Area Unde r the Curv e (p- value )
Vermigli o et al. (2018)	1	HINT (average threshol d)	Speech recognitio n in noise disorder	Self- report	80%	86%	Not applicable	AUC = .86 (p < 0.01)
Vermigli o et al. (2018)	2	Pure- tone threshol d test (bilateral PTA)	Speech recognitio n in noise disorder	Self- report	28%	95%	Not applicable	AUC = .51 (p = 0.94)

#### References

- Beasley, W. C. (1940). Characteristics and distribution of impaired hearing in the population of the United States. J. Acoust. Soc. Am., 12.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D., de Vet, H. C., & Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. The Standards for Reporting of Diagnostic Accuracy Group. *Croat Med J*, 44(5), 639-650.
- Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, *29*(1), 1-7. <u>http://www.ncbi.nlm.nih.gov/pubmed/2310349</u>

- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*, 95(2), 1085-1099. <u>http://www.ncbi.nlm.nih.gov/pubmed/8132902</u>
- Schaette, R., & McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J Neurosci*, *31*(38), 13452-13457. <u>https://doi.org/10.1523/JNEUROSCI.2156-11.2011</u>
- Stilma, W., Rijkenberg, S., Feijen, H. M., Maaskant, J. M., & Endeman, H. (2015). Validation of the Dutch version of the critical-care pain observation tool. *Nurs Crit Care*. <u>https://doi.org/10.1111/nicc.12225</u>
- Vermiglio, A. J. (2008). The American English hearing in noise test. *Int J Audiol*, *47*(6), 386-387. <u>https://doi.org/10.1080/14992020801908251</u>
- Vermiglio, A. J. (2014). On the Clinical Entity in Audiology: (Central) Auditory Processing and Speech Recognition in Noise Disorders. *J Am Acad Audiol*, *25*(9), 904-917. <u>https://doi.org/10.3766/jaaa.25.9.11</u>
- Vermiglio, A. J., Soli, S. D., & Fang, X. (2018). An Argument for Self-Report as a Reference Standard in Audiology. J Am Acad Audiol, 29(3), 206-222. <u>https://doi.org/10.3766/jaaa.16128</u>
- WHO. (2021). World report on hearing (Licence: CC BY-NC-SA 3.0 IGO). World Health Organization.