The reliability and validity of an isiZulu Speech Reception Threshold Test in quiet for hearing and hearing-impaired individuals in KwaZulu-Natal

by

Seema Panday

PNDSEE001

Submitted in fulfilment for the Degree of Doctor of Philosophy (Audiology)

Faculty of Health Sciences

Division of Communication Sciences and Disorders

University of Cape Town

January 2025

Supervisor: Professor Harsha Kathard

Co-supervisor: Professor Wayne J Wilson

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

To my dear children- Bhavesh and Thashil

PLAGIARISM DECLARATION

I, Seema Panday, do hereby declare that:

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the APA 7th edition for citation and referencing for the main thesis. The publications followed APA 6th edition. Each contribution to, and quotation in, this research paper from the work of other people has been attributed, and has been cited and referenced.

This research thesis is my own work and has been submitted to Turnitin and my supervisor has seen the report.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

ABSTRACT

Speech reception threshold testing is part of the initial audiological test battery, and should be in the first language of the individual being tested. A newly developed isiZulu Speech Reception Threshold (SRT) test in KwaZulu-Natal Province, South Africa, required further reliability and validity evidence before it can be used for clinical practice. The aim of this study was therefore twofold: to determine the reliability and the validity of the newly developed isiZulu Speech reception threshold (zSRT) test for individuals with and without hearing loss.

Methods

This study utilised an observational, analytical, repeated measures research design and had two phases. Phase 1 consisted of the reliability, as well as content, convergent, and divergent validity testing components, while Phase 2 consisted of the concurrent validity aspect. These five components were identified for inclusion in a newly developed unified validation framework for this study. Of the 176 participants who were selected through non-probability sampling, 100 had normal hearing and 76 had hearing loss (mild and moderate conductive, moderately severe and severe sensorineural hearing loss).

Results

For the reliability study, the zSRT test was found to have high internal and external consistency (reliability) [ICC values ranged from 0.69 to 0.79]; Bland and Altman analyses showed that the isiZulu SRT test scores differed by no more than 7.5 to 8.7 dB HL between original and repeat assessments. Regarding the content validity, more than 20% of participants rated three of the word recordings below ‘strongly agree’ in the categories of pitch or tone, and for one word recording i.e. /cinga/ below strongly agree in the categories of pitch or tone, clarity or articulation and naturalness or dialect. First language raters proved useful in identifying problematic word recordings that other methods of assessment would have missed. Intraclass correlation coefficient analyses showed zSRT scores were in substantial to very high agreement with the Pure Tone Average (PTA) scores for the normal hearing and hearing loss groups (NH – right ear $ICC_{consistency}$ = 0.78, left ear ICC = 0.67; HL – right ear $ICC_{consistency}$ = 0.97, left ear $ICC_{consistency}$ = 0.95), thereby confirming the convergent validity of the isiZulu SRT test. A mixed model analysis indicated a significant difference between the zSRT and PTA ($p < 0.001$), despite there being no clinical significance (2.9 to 3.0 dB for those with normal hearing, and an average of 2.0 to 3.0 dB for those with hearing loss). For divergent validity, there was a significant difference between the PTA and zSRT and eSRT ($p < 0.001$). However, clinically, there was a 0 – 3dB difference between the PTA and zSRT and 5dB difference between the eSRT and PTA for both ears. The mean psychometric slope (%/dB) at 50% correct perception for all words in the zSRT test was 4.92%/dB for the mild

conductive hearing loss group, 5.26%/dB for the moderate group, 2.85%/dB for the moderately severe sensorineural group, and 2.47%/dB for the severe sensorineural hearing loss group. These slopes were appropriate for the degree of hearing loss observed in each group, confirming the concurrent validity of the zSRT test.

Discussion

The collective examination of each measurement-related element, according to the unified validation framework (reliability; content, convergent, divergent, and concurrent validity), was used to confirm the reliability and validity of the zSRT. The findings suggest the need for multiple levels and methods to be considered to validate speech reception threshold tests to ensure that the test can be used in clinical practice. There is value in including first language community members as raters to establish the content validity of the test, with a broader scientific lens being required when interpreting the SRT and PTA correlation for tonal languages. Contextual factors may impact the overall validation of new speech audiometry tests.

Conclusion

The zSRT test was found to be reliable and valid and can be considered for clinical practice in KwaZulu-Natal Province (notwithstanding the study limitations). The study also offers a new framework for validating speech audiometry tests, which includes both reliability and validity elements, and considers the contextual factors.

ACKNOWLEDGEMENTS

My sincere gratitude and thanks are offered to the *Almighty* for spiritual guidance and protection throughout this journey.

To my supervisors, Professor Harsha Kathard and Professor Wayne Wilson, thank you for your support, guidance, patience, and belief in me as a student. Thank you for your contributions and your professionalism. You have inspired me as an academic. I know this has been a long journey for you, too, but you made me believe that this can be done.

To my family, Vic, Bhavesh, and Thashil- I am grateful for the patience, resilience, and courage you have displayed throughout this journey. Thank you for the laughter, love, encouragement, and the belief that I will get this done.

To my Parents, Mum and Dad- I cannot thank you enough for the prayer, support, and love you have shown me through this process. I am Blessed to have both of you here to see me through this journey.

To my siblings and extended family- Thank you for your encouragement and support.

To the participants- Thank you to all the participants who volunteered their time to participate in this study. *Ngiyabonga*

To Sandile Kgotshane, I am extremely grateful to you for assisting with the data collection and your role as research assistant in this study. Your hard work is appreciated.

To the Oticon Foundation for funding support, which covered the running expenses during the data collection of my study.

To the Dean of UKZN School of Health Sciences and UCDP for the funding support towards relief from my academic duties. I am grateful for the support

To Zamani Dlamini – Thank you for your patience and commitment to the editing of the recording before the start of data collection. Your technical skills are most appreciated.

To Ms Carrin Martin - Thank you for assisting with the scientific editing of my thesis. Your patience and professionalism are most appreciated.

To the Statistical teams- Thank you for your assistance with statistical analysis.

To Prof. Mershen Pillay- Thank you for your contribution during the earlier stages of my thesis development and study. Your wise words will always be appreciated.

To Fikile- Thank you for all your support and for taking care of the children during this journey- *Ngiyabonga*

To my many special friends, colleagues, graduates, and students who have shown interest in my journey, thank you to each one of you. You have inspired me. I am grateful for the many friendships I have developed through this journey. *I am because of who many of you are- Siyabonga*

Finally, to myself, Seema, you have survived a long and tough academic journey. Hard work, belief, resilience and deep faith have seen you through to the end. You will forever be a lifelong learner.

TABLE OF CONTENTS

PLAGIARISM DECLARATION ii
ABSTRACT…… iii
ACKNOWLEDGEMENTS v
LIST OF TABLES viii
LIST OF FIGURES ix
DEFINITION OF TERMS x
LIST OF ABBREVIATIONS xii
CHAPTER 1. INTRODUCTION 1

1.1 Introduction 1
1.2 Background 1

1.1.1 Speech audiometry 2
1.1.2 Speech Reception Threshold testing 2
1.1.3 Current context – SRT test in isiZulu 4

1.3 Problem statement 5
1.4 Research Question 6
1.4 Aims and objectives 6
1.5 Thesis structure 7
CHAPTER 2. SPEECH AUDIOMETRY AND A NEW ISIZULU SRT TEST 9

2.1 Introduction 9
2.2 Speech audiometry 9

2.2.1 The value of speech audiometry 9
2.2.2 Tests in the speech audiometry test battery 10

2.3 Historical development of SRT tests 11

2.3.1 Speech audiometry in South Africa and the challenges it faces 15

2.4 isiZulu SRT test 18

2.4.1 Reasons for an isiZulu test 19

2.5 Conclusion 21
CHAPTER 3. THEORETICAL FRAMEWORK AND METHODOLOGY 22

3.1 Introduction 22
3.2 Theoretical framework 22
3.3 Reliability 26

3.3.1 Test-retest reliability or coefficient of stability 27
3.3.2 Inter-list equivalence or alternate forms of reliability 28
3.3.3 The split-half reliability estimate 29
3.3.4 Intra-and inter-observer reliability 29

3.4 Validity **30**

3.4.1 Content validity *30*

3.4.2 Convergent and divergent validity *32*

3.4.3 Criterion validity *33*

3.4.4 Previous studies investigating the validity of the isiZulu SRT test *36*

3.5 Research paradigm and underlying philosophy **38**

3.5.1 Overview of the Methods *38*

3.6 Ethical considerations **40**

3.7 Conclusion **42**

CHAPTER 4. RESULTS OF STUDY 1: RELIABILITY STUDY 43

Introduction **43**

Conclusion **56**

CHAPTER 5. RESULTS FOR STUDY 2: CONTENT VALIDITY STUDY 57

Introduction **57**

Conclusion **67**

CHAPTER 6. VALIDITY STUDY 3 68

Introduction **68**

Conclusion **80**

6.1 Results of the normative study **82**

6.2 Results for the Divergent validity **83**

CHAPTER 7. DISCUSSION 84

7.1 Introduction **84**

7.2 Discussion **84**

7.2.1 Aim 1: To determine the reliability of the isiZulu test *85*

7.2.2 Aim 2: To determine the construct validation of the newly developed isiZulu SRT test *89*

7.3 Clinical Implications **95**

7.4 Theoretical implications **98**

7.5 Significance of the study **103**

7.6 Limitations **104**

7.7 Recommendations **104**

7.8 Summary **105**

7.9 Conclusion **106**

REFERENCES… 108

APPENDICES… 118

LIST OF TABLES

Table 2.1: Speech audiometry studies in quiet done in languages around the world........................ 12
Table 3.1: Review of studies that correlated the SRT and PTA for newly developed word lists 33
Table 3.2: Review of studies using psychometric function slope to evaluate and select speech audiometry stimuli in quiet 34
Table 3.3: Overview of the Methods for each of the aims and objectives of the study……………..39
Table 4: SRT measurements for parts 1 (internal consistency) and 2 (external consistency) of the study 53
Table 5: ICCs (2,1) and Bland and Altman tests for parts one (internal consistency) and two (external consistency) of the study.. 53
Table 6: Psychometric properties of the 28 disyllabic isiZulu word recordings used in this study 623
Table 7: Friedman analysis of variance analyses comparing raters within each rating category. .. 66
Table 8: PTA and zSRT measurements for right and left ears for participants with normal hearing (NH) and with hearing loss (HL) 76
Table 9: Intraclass correlation coefficient (ICC, 2, 1) results for agreement between PTA and zSRT measures for participants with normal hearing (NH) and with hearing loss (HL)............ 76
Table 10: The linear mixed model analysis results for differences in thresholds by threshold measure (PTA versus zSRT) for participants with normal hearing and participants with hearing loss. 77
Table 11: Psychometric function measures for the 28 isiZulu words for the participants by degree of hearing loss 78
Table 12: PTA and zSRT measurements for right and left ears for participants with normal hearing (NH) according to age group.. 82
Table 13: Mixed model analysis results for PTA average hearing threshold for 0.5, I and 2 kHz and isiZulu SRT (zSRT) and English SRT (e SRT) scores………………………….……83

Please note, the Tables in the inserted publications are numbered sequentially and do not have chapter numbers.

LIST OF FIGURES

Figure 3.1: Validation framework for this study 26
Figure 2: Bland and Altman plots for parts 1 and 2 of the study 54
Figure 3: Psychometric functions for the 28 recorded isiZulu words unadjusted (a) and adjusted (b) for listener hearing thresholds…………………………………………….................63
Figure 4: Percentage of raters strongly agreeing in each rating category for each of the 28 recorded isiZulu word. 77
Figure 5: Psychometric functions for each of the 28 words in the isiZulu SRT test, for each participant group by degree of hearing loss. 78
Figure 7.1: Validation framework for the isiZulu SRT test………………………………..... 85
Figure 7.2: Extended Validation framework for speech audiometry tests in indigenous languages………………………………………………………………………….……100

Please note, the Figures in the inserted publications are numbered sequentially and do not have chapter numbers.

DEFINITION OF TERMS

Concurrent validity	Concurrent validity refers to how well a test can estimate present performance (Maxwell & Satake, 2006).
Construct validity	An overarching type of validity that includes both measurement and design related validity elements and measures accurately the theoretical concept it was intended to measure.
Content validity	Refers to how well a particular test adequately represents the various content domains it was designed to measure (Maxwell & Satake, 2006).
Contextually relevant tools/tests	Those tests or tools available for use with a specific group or population in a setting, their development being based on this specific population (Pascoe, 2011)
Divergent validity	A test that discriminates the newly developed tool or test with other measures that may not accurately measure the construct is known as divergent/discriminant validity
First language	This refers to the language an individual is first exposed to and acquires during early childhood, typically from birth. It is often referred to as the mother tongue or native language.
Hearing impairment	Hearing impairment is an encompassing term that includes all degrees of hearing loss from slight to profound.
Hearing loss	Hearing loss refers to a reduction in the ability to hear sounds at a specific level of difficulty
Home language	This refers to the language most commonly spoken by individuals within their household or immediate family environment. It encompasses the language(s) used for daily communication at home.
Indigenous language	A language that is native to a region or area and spoken by people who originated from that area or land.
isiZulu	isiZulu is one of the 12 official languages in South Africa, being an Nguni language that is spoken by most people in KwaZulu-Natal Province
Psychometric function	The psychometric function of a word is described as its performance-intensity function, or the percentage correct identification of the word as a function of its intensity (Kruger & Kruger, 1997)
Reliability	The extent to which results are repeatable and give the same consistent scores (Joppe, 2000 in Golfshani, 2003; Maxwell & Satake, 2006).
Speech audiometry	Refers to hearing testing procedures that utilise speech stimuli to evaluate auditory function (Lawson 2011).
Speech perception	A complex pattern recognition process where the listener relies on the acoustic, syntactic, semantic cues and overall context (Lyregaard, 1997; Bellis, 2015).
Speech Reception Threshold	Tests used to determine the softest level at which an individual can hear speech 50% of the time (ASHA, 1988).

Validity	Validity refers to how well a test or tool measures what it is intended to measure. (Maxwell& Satake, 2006; Mendel, 2008).
Validation	A process of accumulation of evidence to indicate that a test, product, or system is accurate and intended for its use.
Word Recognition Test	A test that assesses the individual's ability to recognise speech or words when the level is loud enough to obtain a maximum speech recognition score (Mcardle & Hnath-Chrisolm, 2009).

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
BKB (SIN)	Bamford-Kowal-Bench Speech-in-Noise
CID W22	Central Institute of the Deaf Wordlist 22
dB	Decibel
dBHL	Decibel Hearing level
DIN	Digit in Noise test
eSRT	English Speech reception threshold test
HINT	Hearing in Noise Test
HL	Hearing loss
ICC	Intra class correlation
KZN	KwaZulu-Natal
LE	Left ear
NAL-AB	National Acoustic Laboratories Arthur Boothroyd (NAL-AB)
NH	Normal Hearing
PTA	Pure Tone Average
Quick SIN	Quick Sentence in Noise Test
RE	Right ear
SA	South Africa
SIN	Sentence in Noise test
SRT	Speech reception threshold
WIN	Words-in-Noise
ZFLS	isiZulu first language speakers
zSRT	isiZulu Speech reception threshold test

CHAPTER 1. INTRODUCTION

"If you talk to a man in a language he understands, that goes to his head. If you talk to him in his own language, that goes to his heart." Nelson Mandela

1.1 Introduction

South Africa has 12 official languages, nine of which relate to those spoken by the indigenous people of the country, most belonging to the Nguni group of languages. English home-language speakers only account for 9.6% of the population, while the most dominant language, isiZulu, is spoken by 24% of the inhabitants, who belong to the largest Zulu clan in KwaZulu-Natal Province on the east coast of the country. While English and Afrikaans remain the dominant business and education languages, many people do not speak them at home and may therefore not be sufficiently familiar with them to be able to comprehend the nuances that may be required for language use during speech audiometry testing in audiology. This resulted in the development of an isiZulu audiology test by the primary researcher in 2006, this study being the continuation of its validity and reliability testing to establish whether or not it can be used in South Africa.

Chapter 1 describes the background context of the study, which outlines the important constructs of this thesis: speech audiometry, the speech reception threshold test, the clinical value of the SRT tests, and the challenges facing speech reception testing internationally and locally. This is followed by a description of the problem the study intends to address, the aims and objectives, and the outline of the thesis chapters.

1.2 Background

The initial diagnostic evaluation of hearing for both adults and children should include a comprehensive battery of reliable and valid tests that are suitable to assess, quantify and diagnose hearing loss. Diagnosis of a hearing loss requires the audiologist to cross-check each test finding within the test battery, mainly because the initial audiological tests rely on behavioural test measures, including pure tone audiometry, which involves obtaining a behavioural response to a pure tone signal. Pure tone testing is known as the foundation test in audiology (Roeser et al., 2000) and remains the gold standard test to determine type and degree of hearing loss across the speech frequencies (Gelfand, 2009; 2015).

Despite the pure tone test audiometry being important within the initial audiological test battery, it does not provide information about hearing speech and environmental sounds. The auditory system is mainly geared to perceive speech sounds, which vary in level, frequency and timing. Therefore, speech audiometry tests remain a fundamental aspect of the initial audiological test battery (Harris et al., 2007; Ramkissoon et al., 2002).

1.1.1 Speech audiometry

Speech audiometry refers to hearing testing procedures that utilise speech stimuli to evaluate auditory function (Mcardle & Hnath-Chrisolm, 2009). Speech audiometry adds clinical value to the initial audiological test battery by assessing an individual's ability to hear speech sounds used in daily communication. Speech audiometry tests comprise threshold and supra-threshold tests of word recognition, the former including Speech Reception Threshold (SRT) testing and the Speech Detection testing presented in quiet or in noise. The focus of this study is on SRT testing in quiet for adult listeners with and without hearing impairment in the South African context.

In this thesis, hearing impairment and hearing loss are used interchangeably, as supported by WHO in the World Report on Hearing (2021, p. 9). Hearing impairment is an encompassing term that includes all degrees of hearing loss from slight to profound, while hearing loss refers to the reduction in the ability to hear sounds at a specific level of hearing difficulty, which was relevant to this thesis. While it is acknowledged that there are nuanced differences in these terms, there is an overlap, each being used where specifically appropriate.

1.1.2 Speech Reception Threshold testing

SRT tests are used by audiologists to determine the softest level at which an individual can correctly identify speech 50% of the time (ASHA, 1988). These tests are typically administered by presenting familiar words to an individual, either as pre-recorded stimuli (via a compact disc [CD] or similar format) or monitored live voice (via words spoken by the tester and monitored via a sound level meter or audiometer). While SRT test protocols vary, the test words are usually presented in blocks of six, with the stimulus level for each block being decreased until the individual correctly repeats only 50% of the words in a block (McArdle & Wilson, 2006). This level in dB is then deemed to be the SRT test score.

SRT testing is a standard part of an audiologist's initial assessment of hearing for individuals of school age and above (Gelfand, 2009). The clinical value of SRT testing has been extensively reported on in the literature and includes:

1. Quantifying an individual's hearing level for speech (Gelfand, 2009);
2. Cross-checking with the 'gold standard' test for hearing, i.e., pure tone audiometry (Roeser et al., 2000);
3. Using SRT as the reference point for supra-threshold speech testing (Gelfand, 2009);
4. Adding diagnostic and prognostic value to the initial hearing assessment (and to larger diagnostic batteries) (Gelfand, 2009);
5. Evaluating the success of medical, surgical or rehabilitative intervention to restore auditory function (Gelfand, 2009).

Perhaps most importantly, SRT testing serves as one of the primary tests in the initial hearing assessment that directly recognises speech as vital for human communication (Gelfand, 2009; Nissen et al., 2005).

Despite SRT testing proving to be a valuable clinical test, many factors can influence the reliability and validity of the test scores. Factors such as the number of words, type of stimuli, homogeneity of the words, and method and level of the presentation can affect the validity and reliability of the test (Harris et al., 2007). The most important threat to reliability and validity is the listener's familiarity with the test word stimuli (Lyregaard, 1997; Nissen et al., 2005).

To achieve testee familiarity with an SRT (or any speech audiometry) test, the test material should be in the testee's first language (Nissen et al., 2005). First language refers to the language an individual is first exposed to and acquires during early childhood, typically from birth, often being referred to as the mother tongue or native language. Words that have a high frequency of occurrence in a testee's first language are recognised and understood more easily than those with lower frequency of occurrence. Perception of speech is a complex task dependent on both the acoustic properties of the speech signal, and the contextual cues and higher order cognitive processes to recognise and perceive the speech signal. It is well supported in the literature that the test material of speech audiometry tests, specifically for this study on the SRT test, should be in the individual's first language (Harris et al., 2004; Hon-Mensau et al., 2023; Nissen et al., 2005; Nissen et al., 2008). However, most speech audiometry tests, including SRT tests, only exist in the English language due to most initially being developed and standardized in the United States of America and the United Kingdom. The wide application of English-language

tests within contexts where most of the population do not speak English as their first language raises concerns about test validity.

International calls for action to develop linguistically and culturally appropriate speech audiometry tools have resulted in an increased interest in the number of tests developed globally, particularly in countries where English is not spoken as first language. Some examples include tests developed in Tiwi and Walpiri (Plant, 1990), Brazilian Portuguese (Harris et al., 2001), Korean (Harris, Kim, Egget, 2003), Polish (Harris et al., 2004), Mandarin (Nissen et al., 2005), Russian (Harris et al., 2007), Taiwanese Mandarin (Nissen, Harris & Dukes, 2008), Romanian (Olariu et al., 2018) and Ghanian (Hon-Mensau et al., 2023). South Africa has lagged behind in developing and validating speech audiometry tests in any of its 10 indigenous languages, except for some development of Afrikaans word lists, and a South African Dialect English CID word list for SRT developed by Hanekom et al. (2015).

1.1.3 Current context – SRT test in isiZulu

Currently in South Africa, the only available reliable and valid SRT tests (or any speech audiometry tests) are in English, with none being available in isiZulu, despite this being the most spoken of the 12 official South African languages. It is the first language of 23.8% of the South African population, followed by isiXhosa (17.6%) and Sesotho (13.3%), compared to only 9.6% for English. As a result, audiologists in South Africa cannot fully and accurately assess the ability of isiZulu-speaking clients to hear speech. The audiology profession in South Africa has therefore aligned itself with calls to develop and validate culturally and linguistically appropriate test materials for their local populations (Pascoe, 2011).

The process of developing and validating linguistically appropriate SRT test materials for the isiZulu speaking population began in 2006 with the thesis author, Panday (2006), providing guidelines for developing SRT word lists in local South African languages. For an SRT test in isiZulu, Panday (2006) recommended 28 familiar, low tone, phonetically dissimilar, bisyllabic verb imperatives, which were shown to be homogenous with respect to audibility with an average psychometric function slope of 6%/dB. These 28 words were used to establish SRT scores using the isiZulu word list on a minimal sample of 30 healthy young adults (18–30 years). There was a strong correlation between the pure tone average (PTA) and the isiZulu SRT score (r= 0.76, $p < 0.05$) for this group of normal hearing individuals. Panday (2006) and Panday et al. (2007, 2009) focused on the development and identification of the test words and

further refined word selection for homogeneity of audibility. As part of the process towards developing the test, certain aspects of reliability and validity were also focused on in the 2006 study, including intra-rater reliability, and aspects of content validity and convergent validity.

1.3 Problem statement

Speech audiometry, specifically SRT testing, is an important tool for assessing hearing function and assessing an individual's ability to recognize and comprehend spoken language, providing a more valid measure of auditory function in real-world listening environments compared to pure tone audiometry alone. Additionally, for the results to be valid and meaningful, the test must be conducted in the listener's first language to avoid linguistic bias that could misrepresent true hearing ability (Nissen et al., 2005). In South Africa, the majority of speech audiometry tests are based on English-language standards that were developed in the United States of America and the United Kingdom. This poses a challenge in a multilingual country where only 9.6% of the population speaks English as their first language, and the majority of individuals communicate primarily in one of the country's indigenous African languages. While a limited number of tests have been developed in Afrikaans, no standardized speech audiometric tests exist for most of the indigenous languages spoken in South Africa.

The absence of linguistically matched speech audiometry tests in South Africa presents a clinical gap, as research has shown that speech audiometry tests, particularly SRT tests, must be administered in the patient's first language to yield accurate and reliable results (Nissen et al., 2005). In response to this need, Panday (2006) developed a new isiZulu SRT (zSRT) test that was designed to assess the hearing thresholds of isiZulu-speaking individuals, the largest linguistic group in South Africa. The test, consisting of 28 bisyllabic, low-tone verb imperatives, was carefully constructed to align with the unique phonetic and linguistic features of isiZulu. Despite preliminary reliability and validity testing, the test lacks comprehensive and rigorous evidence supporting its psychometric properties, particularly in the context of clinical practice.

For the zSRT test to be considered suitable for widespread clinical use, its reliability (consistency of test results over time) and validity (accuracy in measuring speech reception) must be thoroughly evaluated. Despite the growing recognition of the importance of language-specific audiometry, no study has systematically examined the reliability and validity of Panday's zSRT test in the context of the diverse population of KwaZulu-Natal (KZN) Province,

where the test is most likely to be used. This study therefore sought to address this research gap by determining the reliability and validity of the zSRT test for hearing and hearing-impaired individuals in KZN utilising an integrated validation framework.

The findings from this study will contribute to the development of culturally appropriate, reliable and valid speech reception threshold tests, which are essential for equitable hearing assessments in South Africa. Reliability and validity are interrelated (Maxwell & Satake, 2006; Mendel, 2008), and their full realisation therefore can only occur within an integrated, unified validation framework, which will be used to meet this study's aims and objectives.

1.4 Research Question

The research question for this study was: Is the newly developed isiZulu SRT test for adults sufficiently reliable and valid for clinical use in KwaZulu-Natal (KZN) Province in South Africa?

1.4 Aims and objectives

The study was directed by two aims, four sub-aims and eight objectives.

Aim 1: To determine the reliability of the isiZulu SRT test in quiet for isiZulu speaking adults in KZN.

Objective 1.1: To determine the test-retest reliability (external consistency) of the isiZulu SRT test.

Objective 1.2: To determine the internal consistency (spilt-half reliability) of the isiZulu SRT test.

Aim 2: To determine the construct validity of the isiZulu SRT test in quiet for isiZulu-speaking hearing and hearing-impaired adults in KZN

- **Sub-aim 2.1:** To determine the content validity of the isiZulu SRT test.

  Objective 2.1.1: To determine to what extent the SRT test recording represents the natural dialect, intonation and tone of the isiZulu language spoken in KZN.

- **Sub-aim 2.2:** To determine the convergent validity of the isiZulu SRT test.

  Objective 2.2.1: To establish normative data for the isiZulu SRT test for isiZulu-speaking adults (18–60 years).

  Objective 2.2.2: To determine any correlation between the isiZulu SRT score and PTA for hearing and hearing-impaired adults (18–60 years).

- **Sub-aim 2.3:** To determine the divergent validity of the isiZulu SRT test

  Objective 2.3.1: To determine if there is a difference among the isiZulu SRT and English (CID W2) SRT scores and PTA for hearing and hearing-impaired adults.

- **Sub Aim 2.4:** To determine the concurrent validity of the isiZulu SRT test.

  Objective 2.4.1: To determine the percentage correct identification of the isiZulu SRT words at specified intensity levels for participants with hearing impairment.

  Objective 2.4.2: To determine if there is a difference in the performance-intensity functions of the isiZulu SRT words between hearing and hearing-impaired adult groups.

1.5 Thesis structure

This thesis is presented in a hybrid format, with the results being provided in three chapters as published studies. *I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include 3 publication(s) in my thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publications (refer to Appendix K for declaration and list of publications included).*

Chapter 2. Speech audiometry and a new isiZulu SRT test for South Africa: The chapter provides a more detailed discussion of the constructs introduced in Chapter 1, including a definition of speech audiometry, its clinical value, the status of speech audiometry in South Africa and the challenges this presents, the development of speech audiometry tests for South Africa, and an isiZulu SRT with a description of its development until the start of this study.

Chapter 3. Theoretical framework and methodology: This chapter describes the theoretical framework and overall methodology of the research reported in the thesis. A unified approach of validity is described, as well as theoretically how the various measurement related elements of the thesis contribute to an overall validation framework. It provides the various levels of reliability and validity testing required for a test to be considered suitable for clinical practice, after which a validation framework is suggested. This chapter also provides the research paradigm of the thesis. As this is a hybrid submission, the method for each objective is described in each of the publications presented in Chapters 4, 5 and 6. However, an overview of the methods related to each of the objectives and framework of the study is presented in table format for ease of reference.

Chapter 4. Results for Study 1 (Reliability study) This chapter presents Study 1, the reliability study, addressing Objective 1.1, which focuses on the test-retest reliability (external consistency of the test), and Objective 1.2, which focuses on the split half reliability (internal consistency of the test). A paper titled "The Internal and External Consistency of the Newly Developed Word Speech Reception Threshold Test for isiZulu Speakers with Normal Hearing Sensitivity" was published in the South African Journal of Communication Disorders in 2018. This chapter consists of the paper as it was published, with an expansion of the methods section and minor editorial changes to the layout of tables and figures.

Chapter 5. Results for Study 2: (Content Validity Study) This chapter presents Study 2, the content validity study, addressing Objective 2.1.1. A paper titled "First Language Raters' Opinions when Validating Word Recordings for a Newly Developed Speech Reception Threshold Test" was published in the South African Journal of Communication Disorders in 2018. This chapter consists of the paper as it was published, with an expansion of the methods section and minor editorial changes to the layout of tables and figures.

Chapter 6. Results for Study 3: (Convergent, Divergent and Concurrent Validity). This chapter presents Study 3, the convergent and concurrent validity study, addressing Objectives 2.2.1, 2.2.2, 2.3.1, 2.4.1, and 2.4.2. It is presented as a publication "The Validity of an isiZulu Speech Reception Threshold Test for Use with Adult isiZulu Speakers" that was published in the South African Journal of Communication Disorders in 2020. The chapter includes the paper as it was published, with an expansion of the methods section, and provides minor editorial changes to the layout of tables and figures. This chapter also provides the results of the normative study (addressing Objective 2.2.1) and the divergent validity study (addressing Objective 2.3.1).

Chapter 7. **Discussion and Conclusion** This chapter discusses the research presented in the thesis and describes to what extent the zSRT test is valid and reliable for clinical use in the adult isiZulu speaking population of KZN, South Africa. The significance of the findings for regional and international contexts is also discussed, as are the thesis limitations and the recommendations made for policy, practice, and research more broadly.

CHAPTER 2. SPEECH AUDIOMETRY AND A NEW ISIZULU SRT TEST

2.1 Introduction

This chapter begins with a background to speech audiometry, followed by a description of SRT testing and SRT tests in South Africa, the rationale for using these tests in audiology, and the challenges faced in South Africa regarding speech audiometry. The chapter also provides the background for the SRT test in isiZulu developed by Panday (2006).

2.2 Speech audiometry

Speech audiometry consists of a battery of tests or procedures in audiology that use speech stimuli, such as phonemes, words, sentences or digits, to evoke a behavioural response from a listener, including, but not limited to, indicating if the stimulus was heard, repeating the stimulus as heard, or pointing to a written version of the stimulus or picture representing the stimulus (Lyregaard, 1997). Speech audiometry remains an important component of the initial audiological test battery, others being case history, otoscopic examination, pure tone audiometry and acoustic immittance testing (Ma et al., 2013).

2.2.1 The value of speech audiometry

Lyregaard (1997) argues that one of the most important human abilities is to understand and comprehend speech and language as a means of communication. Assessing this aspect of the auditory system requires audiological tests that are standardised, sensitive, reliable and valid. Speech audiometry tests remain valuable components within the audiological test battery, because the auditory system is sensitive to the perception of speech (Ramkissoon, Estis & Flagge, 2014). Although the conventional pure-tone test is regarded as the gold standard in the initial audiological test battery (Roeser, et al., 2000), it falls short of providing information about the hearing of everyday communication. Speech signals used for speech audiometry tests are more complex and represent the sounds of everyday human communication.

Speech audiometry is useful, as it complements the conventional air and bone conduction pure tone tests. Ma et al. (2013) summarise the key purposes of speech audiometry tests as follows:

- To diagnose the site-of-lesion for auditory disorders.
- To evaluate central auditory processing abilities.

- To determine the success of rehabilitative and intervention methods in audiology such as hearing aids and cochlear implants.

2.2.2 Tests in the speech audiometry test battery

Many tests comprise the speech audiometry test battery, these being grouped into those that assess the sensitivity (audibility aspect) and clarity (distortion aspect) of speech (McArdle & Hnath-Chisolm, 2009). To assess the sensitivity of speech, audiologists use threshold measure tests, such as speech detection threshold (SDT) tests and the speech reception threshold (SRT) tests. To assess the clarity of speech, audiologists use supra-threshold measure tests, such as word recognition (WR) tests or speech recognition tests. Each test has specific clinical value within the initial audiological test battery. SRT tests were historically conducted in quiet as part of the initial audiological test battery, but it is now recommended that SRT and suprathreshold testing are conducted both in quiet and in noise to better consider the common complaint of difficulty understanding speech in background noise (Smits, 2013).

Early speech-in-noise tests (e.g., Plomp & Mimphen, 1979) used English sentences in background noise. Subsequent speech-in-noise tests have used sentences or words-in-noise including tests, such as the Sentence in Noise (SIN) test, Hearing in Noise Test (HINT), Quick SIN, Bamford-Kowal-Bench Speech-in-Noise (BKB SIN) test; Words-in-Noise (WIN); Connected Speech Test (CST), Sentence in Noise test in Afrikaans, and the Digits in Noise (DIN) test (Thakor, 2020). Although the speech-in-noise tests have expanded and are well-supported in the literature, Smits (2013) notes these tests are often not included in the initial test battery. This lack of inclusion has also been noted in South Africa (Thakor, 2020), although there is a Sentence-in-Noise test being available in Afrikaans and sentence- and digit-in-noise tests being widely used for hearing screening (Smits, 2013). There have also been successful recordings of the tests, such as the Hearing in Noise (HINT) test, into several languages, including French-Canadian, Swedish, Brazilian Portuguese, and Mandarin (Melo et al., 2017; Vaillancourt et al., 2005; Wong et al., 2007). While the clinical value and the well-established ecological validity of the speech in noise and digit in noise tests are acknowledged, this study focused on a test in quiet as an initial case example for validating a test in the local languages in South Africa as opposed to a speech test in noise based on several contextually grounded reasons. These include:

- Clinical relevance and contextual appropriateness: the use of SRT testing in quiet remains the standard clinical procedure in audiological assessment in KZN. This choice was driven by

the current diagnostic protocols, availability of audiological infrastructure and clinical training practices in the region.

- Existing isiZulu SRT test: the isiZulu SRT test used in this study was developed in a previous study. The test materials have undergone extensive linguistic, cultural and audiometric evaluation, ensuring their appropriateness for the target population.
- The SRT in quiet: this test is not only a widely used tool in clinical audiology internationally, but represents the foundational step in speech audiometry test development. A review of the literature indicates that SRT tests in quiet are commonly developed first, due to their methodological simplicity, controlled acoustic environments, and importance in establishing normative and threshold data. These foundational measures are essential precursors to the development of more complex speech-in-noise measures. The isiZulu SRT test thus lays the groundwork for future adaptation and expansion into speech-in-noise paradigms.

2.3 Historical development of SRT tests

The first notable work using spondaic words (where a spondaic word, or spondee, is a bisyllabic word spoken with equal stress on each syllable) to assess SRT in quiet was the Psychological Acoustic Laboratory (PAL) auditory test No. 9 (Hudgins et al., 1947) in British English. This was soon followed by the Central Institute for the Deaf (CID) wordlists 1 and 2 in American (USA) English (Hirsh, 1952). Table 2.1 provides examples of bisyllabic and trisyllabic word tests subsequently developed for SRT testing in quiet in adults with first languages other than English.

Researchers investigating word choice for SRT testing in quiet have proposed specific criteria for selecting words to use in the test. These criteria include familiarity, which refers to the most commonly and frequently used words in the language (Hudgins et al., 1947); homogeneity of audibility, which refers to the "ease at which the words are understood when spoken at a constant level" (Silman & Silverman, 1991 p. 40); phonetic dissimilarity, which means that the word list must have varying combinations of consonants and vowels and should not be too easy for listeners to identify; and normal sampling of sounds in the English language. These criteria have remained over the years, with familiarity and homogeneity of audibility receiving increasing prominence and acceptance.

Table 2.1: Speech audiometry studies in quiet completed in various languages

Study/ Researchers	Language	Findings in normally hearing adults (unless stated otherwise)
Plant (1990)	Walpiri, Tiwi Languages Aboriginal – SRT	Speech materials developed in both languages using bisyllabic nouns.
Harris et al. (2003)	Korean	Bisyllabic SRT words found to have a mean slope at 50% correct of 10.35/dB for the male speaker and 9.0%/dB for the female speaker.
Harris et al. (2004)	Polish	Bisyllabic SRT words found to have a mean slope at 50% correct of 10.1%/dB for the male speaker and 9.8%/dB for the female speaker.
Nissen et al. (2005)	Mandarin	24 trisyllabic Mandarin words for SRT were found to have a mean slope at 50% correct of 11.3%/dB for the male speaker and 12.1%/dB the female speaker.
Christensen (1995), as cited in Nissen et al. (2005)	Spanish	Trisyllabic words.
Harris et al. (2007)	Russian	25 bisyllabic words found to have a mean slope at 50% correct of 12%/dB for the male speaker and 9.9%/.dB for the female speaker.
Wang et al. (2007)	Mandarin (Chinese)	Disyllabic words found to have a mean slope at 50% correct of 2.7%/dB for hearing-impaired adults and 5.6%/dB for normally hearing adults.
Nissen et al. (2008)	Taiwan Mandarin	Trisyllabic words found to have a mean slope at 50% correct of 8.3%/dB for the male speaker and 6.8%/dB for the female speaker. Slopes were reported to be steeper than English words due to differences in linguistic structure.
Harris, Mcpherson, Hanson & Egget, (2017)	Vietnamese	48 bisyllabic words found to have a mean slope at 50% correct of 9%/dB. Further testing is recommended with hearing-impaired populations.
Han et al. (2011)	Korean	36 bisyllabic words. Mean SRT score was 23.44 dB SPL, 3.44dB better than mean SRT score for English words (spondees). Some reasons provided for the difference between the two languages: Korean speech sounds at hearing level may be difficult to understand due to a higher concentration of energy in the low frequency bands compared to the English speech spectrum. English words were spondaic words (equally stressed words) while Korean words have the stress on the first syllable.
Olariu, 2018	Romanian	Strong correlations between SRT scores and pure tone results.
Hon Mensah et al. 2023	Ghanian	25 words found to have slopes at 50 % correct greater than 7%/dB were selected and recorded for SRT testing.

Familiarity is achieved by selecting those words or stimuli that are common and frequently used by the communities that speak the language as their native or first language. Words or stimuli having a higher frequency of occurrence in the target population (familiar words) are more easily recognised by people in that population than are words with lower frequencies of

occurrence (Craig, 1997). It is also argued that listeners rely on both bottom-up (decoding of the acoustic signal such as frequency, level and timing of sounds) and top-down (use of higher order cognitive functions such as prosodic, semantic, lexical knowledge of the language and the context) processing to understand and recognise speech (Bellis, 2015; Lyregaard, 1997; Medwedsky, 2002). Furthermore, first language speakers are likely to perform much better than second language speakers when presented with speech materials in their native language. Takayangi et al. (2002) measured lexical difficulty of word recognition tasks for first and second language speakers of English, with hearing and hearing-impaired groups. Their study found that that the second language English speakers in both the hearing and hearing-impaired groups required greater intensity for the equal intelligibility of the words.

The homogeneity of audibility refers to the "ease at which the words are understood when spoken at a constant level of intensity" (Silman & Silverman, 1991 p. 40), and is often achieved by selecting only those words that are heard at the same intensity level, or by recording words in such way that they are all heard at the same intensity level. Homogeneity of audibility can be achieved by determining how well a list of words/ test items can be heard in relation to the intensity function. The common method to select homogenous words is by recording the percentage of correctly heard words at varying intensity levels, which is usually a psychometric function curve. The psychometric function of a word is described as its performance-intensity function, or the percentage correct identification of the word as a function of its intensity (Kruger & Kruger, 1997).

For the words in an SRT test, these functions should be homogenous in audibility and have steep psychometric function slopes. The psychometric function illustrates how well the speech signal can be correctly identified as a function of intensity levels. Critical parameters of the psychometric function curve may be used to identify the most homogenous words. Two important factors include the threshold (50% point) and the slope of the curve. The 50% performance level determines the location of the function, whereas the slope determines the shape of the function (i.e. 20% - 80% correct level) (Shi & Zaki, 2014). A steep psychometric curve means that a small increase in signal in presentation level would lead to a large increase in intelligibility, and that if the slope is relatively shallow, the same presentation level would lead to a smaller perceptual improvement (MacPherson & Akeroyd, 2014).

A steep slope is preferred, as it provides more precise estimations of SRT (Hudgins et al., 1947) and greater sensitivity to changes in a listener's ability to understand speech (Theunissen et al., 2009). From the review presented in Table 2.1, slopes for existing SRT tests range from 7.2%/dB to 12.1%/dB for bisyllabic or trisyllabic words in Italian, Korean, Mandarin, Polish, Russian, Spanish, Taiwanese Mandarin and Vietnamese, and more recently in Fante, which is an African Ghanian language. These findings suggest that steep and favourable psychometric function curves are a possibility in languages where spondees words do not exist. These findings also suggest and confirm that if the stimuli or signal is familiar to the individual and homogenous with regards to audibility, the psychometric functions of individual words would be relatively steep.

Table 2.1's review of research on SRT tests in quiet in languages other than English shows that researchers had to adapt the original criteria for test development to suit the language. One adaptation in particular was word type. Original criteria for SRT tests in English recommended the use of bisyllabic, equally stressed (spondee) words. In the English language, spondee words are usually bisyllabic compound nouns (Hodgeson, 1980), such word types seldom existing in languages other than English. Hodgeson (1980) also argued that spondees do not exist naturally in the English language, but rather exist as bisyllabic words that can carry equal stress on both syllables without altering the meaning of the words (e.g. baseball, hotdog, and airplane). This allowance for variation in the stress pattern without changing meaning is not common in all languages and does not occur in tonal languages, such as Mandarin, Korean, Thai and African languages, such as isiZulu, where changes in the stress of a syllable change the meaning of the word. The lack of equally stressed words in such languages resulted in the use of bisyllabic words with trochees (stress on the first syllable) and iambs (stress on the last syllable) stress, as well as trisyllabic words with varying emphases. The outcomes of these studies have been favourable when the homogeneity of audibility and performance intensity functions was described (Table 2.1).

The review of the non-English language SRT test literature also shows a focus on normal hearing participants only. It is known that if tests are to be used within the clinical population, individuals with hearing loss may show variability in the psychometric pattern (MacPherson & Akeroyd, 2014). There is a paucity of research on the psychometric functions of words/speech signals in quiet for individuals with hearing loss. For the few studies that have been documented, the patterns that emerge indicate that psychometric function curves were shallow

in comparison to the curves for participants with normal hearing and these were typical for the type and degree of hearing loss. There is often considerable variability in the curves (Wang et al., 2007)

Different types and degrees of hearing loss are known to affect individuals' speech recognition and perception (Wilson & Carter, 2001). Certain characteristics of the hearing loss itself may impact an individual's speech perceptual abilities. With a decrease in the audibility of the signal, individuals with hearing loss have difficulty in extracting information (acoustic) from speech signals to aid their hearing and understanding of speech. The temporal and spectral resolution abilities are affected, especially in typical sensorineural hearing loss and with older adults (Andrade et al., 2023).

Usually, the frequencies of 500Hz, 1000Hz and 2000Hz are most sensitive to speech sounds, while vowel and consonant sounds have varying spectral energy. It is known that sounds are usually intense in lower frequencies, and consonants have more intense energy in the high frequencies. Due to the spectral characteristics of these sounds, individuals with hearing loss may have difficulty with frequency selectivity and temporal resolution, and with extracting the cues for speech understanding. Therefore, when new tests are developed, such as the isiZulu SRT, investigating the validity of these new tests on individuals with hearing loss is important. The variation in the energy content, speech spectrum, and acoustic properties of vowels and consonants in the different languages may also impact the speech perception of individuals with hearing loss. It has already been shown in Panday's (2006) study that the acoustic properties of isiZulu verbs are typically low-high tone words, e.g. */banga/*. Listening to these words with a hearing loss may be worth investigating for the reasons already discussed above. Validity outcomes for hearing loss groups compared to their hearing counterparts may not be the same.

2.3.1 Speech audiometry in South Africa and the challenges it faces

South Africa has 12 official languages, with most of its population speaking an African language as their first language (Pascoe, 2011). Within this context, the audiology profession (along with many others) has recognised the need for culturally and linguistically relevant resources (Pascoe, 2011). Although this need is widely acknowledged, attempts to address it remain few in number. This is due, in part, to the audiology profession (and many others) still recovering and transforming from the historical past of inequality and lack of service provision

to the majority of South Africans, the need to intellectualise and implement services in all languages.

In 2019, the Health Professions Council of South Africa's (HPCSA) Board of Speech, Language and Hearing, published a guideline document for practitioners that advocates five important principles when working with clients from culturally and linguistically diverse backgrounds. The document articulates the need to prioritise services in the client's first language, and to use assessment tools that take their culture and language into account to ensure a culturally fair assessment.

Locally, the English language CID W1 and W2 wordlists and their subsequent adaptations to suit the local English dialect (Hanekom et al., 2015), and the development of Afrikaans speech audiometry tests, remain the only options available for South African audiologists when assessing the population's hearing of speech. This situation has remained, despite democracy having been in place in South Africa for 30 years, with most of the South African population speaking English as a second, third or fourth language, and the knowledge that SRT testing is best conducted using test materials in the listener's first language (Khoza-Shangase & Mophosho, 2018; Nissen et al., 2005; Panday et al., 2007; Pascoe, 2011).

Two approaches have been taken to address the need for linguistically appropriate test materials in speech audiometry, these being to use non-South African speech audiometry tests for the South African population, and to develop new speech audiometry tests. With regards to the first approach, South African speakers were tested using the National Acoustic Laboratories (NAL) word list (Wilson, Jones, & Fridjhon, 1998) and the Central Institute of the Deaf Wordlist 22 (CID W22) in American English (Wilson & Moodley, 2000). It was found that first language speakers of South African English with normal hearing thresholds performed well on non-South African English speech recognition tests at supra-threshold levels but poorly at threshold levels. These findings suggest that adapting or using non-South African language speech audiometry tests for South African populations is not a suitable option.

Regarding the development of new test materials for South African speakers, Hanekom et al. (2015) developed a South African dialect English spondaic word list. Although the findings of the study are promising, the study focused on normal hearing populations, with much of the attention for this test development being on selecting English words for the South African

dialect. It’s widescale application to the clinical hearing loss population would require further reliability and validity testing, mainly because speech understanding can be influenced by varying degrees and types of hearing loss.

Furthermore, the research findings in this study are only applicable to the population of the sample selected for the research, i.e. second language speakers of English, who only account for a small percentage of the South African population. The recent general household survey in South Africa released in May 2019, and the census 2022, confirmed that English is only spoken as a first language by 8.1% of the population compared to 22% who speak isiZulu. In addition, outside the home, English is only spoken by 16% of the population, compared to 25% who speak isiZulu (STATS SA, 2019). This suggests that a large percentage of isiZulu speakers use isiZulu as a first language both in and outside the household, and their familiarity with the language would be far better than with English. Although an English spondaic word list exists in South Africa, the lack of reliability and validity data, and in view of the limited sampling, suggests that a first language isiZulu test for SRT is required for wider usage in South Africa, specifically in KZN, where the majority live or have family homes that they return to.

Ramma and Khoza (2008) utilised an informal list of Tswana words for an SRT test as an alternative to the CID W1 English test. In their study, young, healthy, second language English speakers between 18 and 25 years old were tested with the informal Tswana bisyllabic word list, the CID W1 list and a paired digit test. The findings obtained suggest that the Tswana word list had a stronger correlation with the PTA of 0.61 when the Pearson rank test was utilised compared to English and digit pairs. This was similar to the Hanekom et al. (2015) study, which was limited to second language speakers of English only and focused on young, healthy, university students and staff whose literacy levels would not be comparable to the general population. The findings and recommendations in this study were also consistent with both national and international views regarding the need to develop locally specific word lists in the indigenous languages in South Africa. Added to this, these authors also recommended norming such tests on the local populations with and without hearing loss. Normative data for newly developed tests should exist before their implementation within the clinical population and must be normed on the population for whom it is designed (Wyatt, 2012). However, various socio-economic and demographic factors, as well as the geographical locations of participants, can also impact the study outcomes for the development of normative data.

These factors have also been shown to impact research findings relating to speech perception tests. If a test is being developed for a dominant language in a community, then recruitment of participants who only speak this language should be easily achieved; however, this situation is not always possible in multilingual communities. In African countries where individuals are exposed to several languages and speak and read more than one, the selection of participants becomes a difficult and complex issue. These factors have continued to challenge researchers locally and internationally, and may have stalled efforts to develop such tests. This may call for broader and alternate views for test development, standardisation and validation in multilingual contexts, with language proficiency needing to be considered in the standardised samples.

Given the complexity of test development and validation, translating tools from English into indigenous languages is often the primary strategy to address the limitations of available resources. While some of the research related to translation of tools may have shown promising results, there is also compelling evidence that translating tools from one language to another may not be appropriate, mainly because languages differ in structure, sound systems, dialect, word corpus, semantics and syntax. Pascoe (2011) argued that this approach to translate tools directly from languages such as English to indigenous languages is not usually recommended. This research provides examples of when such tools were used with indigenous language speakers, whose test performance was less than optimal. Wyatt (2012) recommended that if translated tools are utilised, they should meet the guidelines recommended by the International Test Commission. Where tools do not exist for a participant language group, it is recommended that they should be developed for that language. Similarly, in South Africa, the recommended approach to meet the need of culturally and linguistically relevant tools would be to develop them to suit the language and population for whom it is meant (HPCSA, 2019), which has implications for the isiZulu test, and its subsequent development and validation.

2.4 isiZulu SRT test

The isiZulu SRT test development began as of part a master’s project that was completed in 2006 by this thesis’s author (Panday). With the paucity of linguistically matched SRT word lists in South Africa, that project aimed to develop and assess the isiZulu SRT test on a sample of normal hearing individuals. The researcher focused on collecting the test items, rating the words for criteria that matched the linguistic and audiological criteria, testing them for psychometric properties and utilising the SRT word list as part of a basic test battery for normal

hearing, healthy young adults. The development and assessment of the word list were provided in two publications by Panday et al. (2007; 2009).

As part of Panday's 2006 study, two tertiary level educators and two isiZulu speaking language interpreters identified 131 commonly used isiZulu words for potential inclusion in an isiZulu SRT test, of which 124 were regarded as bisyllabic verbs. Five linguists (each having a master's degree in the linguistics of African languages) rated 58 of the verbs as being phonetically dissimilar, low in tone and familiar for inclusion in the isiZulu SRT test (Panday et al., 2007). Recorded versions of the remaining 58 bisyllabic words (spoken by an male adult, first language isiZulu speaker) were assessed for homogeneity of audibility by playing the words at six levels of intensity to 30 isiZulu, first language speaking adults (aged 18 to 25 years) who had normal hearing (Panday et al., 2009). Homogeneity of audibility was established by examining the psychometric functions (also known as the performance-intensity functions or curves) of the 58-word recordings, which indicated how well they were each correctly identified as a function of intensity level. Of the 58 words, 28 met the criteria of having a mean slope at 50% intelligibility within 1 SD of the group mean of 5.98%/dB. Lastly, the analysis of the prosodic features and pitch contours of the 28 word recordings showed that they conformed to the prosodic pattern that is apparent within the isiZulu linguistic structure (Panday et al., 2009).

2.4.1 Reasons for an isiZulu test

South Africa is a multilingual and multicultural society with 12 official languages, isiZulu being recognised as one of them after the transition to democracy in 1994. In addition, isiZulu, which is part of the Nguni family of languages in South Africa, is spoken by 22.7% of the population, and understood by approximately 50% of the population nationally, and is the most widely spoken and understood language in the province of KZN. IsiZulu is also spoken in parts of Gauteng Province neighbouring KZN, with migrants having taken isiZulu to the adjacent countries of Zimbabwe and Mozambique.

isiZulu is deeply intertwined with the culture of the Zulu people of KZN, South Africa (Rudwick, 2008). While the indigenous language of Afrikaans was recognised and intellectualised due to it having been spoken by the ruling class of White settlers before democracy in 1994, the colonial and apartheid laws and practices prevented the acknowledgment and upliftment of African indigenous languages. Since 1994 (the end of apartheid), there has been an increase in the use of African languages in the print and electronic

media and courts, and policies promoting African language use in the basic and higher education sectors.

Despite the laws and policies on the formal use of isiZulu, of importance to this thesis is that the language is a key carrier and medium for the culture of the people, mainly at grassroot levels. Rudwick (2008) explains how the use of isiZulu language among young individuals, particularly in rural households, is seen as honouring and respecting their culture. isiZulu continues to be a language of choice for those who speak it as a home language, despite the effects of language shifting, language borrowing and code switching within the central geographical urban regions of KZN due to acculturation, education and increased multilingual communication (Rudwick, 2008).

The isiZulu linguistic structure is complex, and differs considerably from Germanic languages such as English. Of particular relevance to the present thesis are two features of isiZulu: tonality and the absence of spondaic words. First, isiZulu is a tonal language (Rycroft & Ngcobo, 1979), with variations in pitch influencing word meaning (similar to Mandarin and Cantonese). Tone variation usually occurs at the syllable level, with two words with the same phonological structure differing in meaning through the use of three main tones: low, high and falling (Van de Merwe & Le Roux, 2014). Second, isiZulu does not contain spondaic words (bisyllabic words with equal stress on each syllable) of the type recommended for English-language SRT test materials. In isiZulu, stress or prominence mainly falls on the penultimate syllable of words, which often occurs when the second syllable is shortened (Cope, 1982).

A further feature of isiZulu is its complex consonant system and use of unique click consonants. The click sounds are divided into three positions of articulation: dental alveolar, alveolar and lateral. isiZulu also has breathy consonant phonemes, known as depressor consonants, which can affect a words' acoustic properties by lowering its overall pitch. In contrast to its complex consonant system, isiZulu has a simpler vowel system consisting mainly of five vowels with sounds similar to English language vowels.

The features of isiZulu discussed above were considered by the present thesis author (Panday) when beginning the process of developing a contextually relevant isiZulu SRT test for South Africa (Panday, 2006; Panday et al., 2007). Particular attention was paid to the linguistic structure of isiZulu and how it aligns with the requirements for SRT test materials. For an SRT

test in isiZulu, Panday (2006) recommended 28 familiar, low tone, phonetically dissimilar, bisyllabic verb imperatives that were homogenous with respect to audibility (each showing an average psychometric function slope of 6%/dB). These 28 words were used to establish SRT scores using the isiZulu word list on a minimal sample of 26 healthy young adults (18–30 years). There was a strong correlation between the pure tone average and the isiZulu SRT score (r= 0.76, $p<0.05$) for this group of normal-hearing individuals. As part of the process towards developing the isiZulu SRT test, Panday (2006) focused only on intra-rater reliability and convergent validity.

While significant strides were made towards the development of the first isiZulu SRT test in quiet in KZN, the resulting word list showed variability in the performance intensity curves of the words after the homogeneity of audiology was investigated in the Panday 2006 study. To resolve this concern prior to starting the present thesis, further adjustments were made to the recording of the words in the isiZulu wordlist. Each isiZulu word was adjusted in level (Panday et al., 2018) to ensure its 50% correct perception score occurred at the mean pure tone average (2.8 dB HL) of the 30 normally hearing isiZulu first language speaking adults (mean age 21.5 years) who participated in the Panday et al. (2009) study. Once adjustments were made to the recording of the words that showed the most variability, the zSRT test was ready for further consideration of its reliability and validity in normal and clinical, first language isiZulu speaking adult populations.

2.5 Conclusion

This chapter provided a background to speech audiometry and the rationale for using SRT testing in the initial audiological test battery. The lack of a culturally and linguistically relevant SRT test for isiZulu speakers in KZN, South Africa, was discussed, as was the preliminary work of the current thesis author (Panday) to develop such a test. The chapter identified the need to complete the development of the zSRT test to confirm its reliability and validity in normal and clinical isiZulu speaking adult populations in KZN, South Africa.

CHAPTER 3. THEORETICAL FRAMEWORK AND METHODOLOGY

3.1 Introduction

This chapter begins with a historical overview of validity theories in fields related to audiology, and a discussion of the unified theory of construct validation and its suitability to the present thesis's development of an isiZulu SRT test. As part of the overall methodology of the thesis, the research paradigm and philosophy are presented, followed by a review of reliability and validity. As this is a hybrid thesis submission, the methods used in each study within the thesis are described in the methods sections in Chapters 4, 5 and 6. For ease of reference, this chapter provides an overview of these methods in tabular format.

3.2 Theoretical framework

The concept of reliability and validity is not new to the field of audiology where researchers have sought to develop valid and reliable tests of hearing soon after the middle of the 20th century. This long history of test development has seen a wide range of approaches, designs, methods and theoretical frameworks.

Historically, the debate on the value of validity theory in audiology has roots in similar debates in education and psychology, with research in communication sciences and disorders aligning with the recommendations of American Psychological Association and educational fields (McCauley & Swisher, 1984). Hathcoat (2013) provides an overview of the historical evolution of validity theory in the field of educational and psychological testing that began with the notion of 'testing', scoring and quantifying levels of academic achievement and intelligence quotient attributes, which resulted in increasing debates about the validity of these tests. Hathcoat (2013) indicates that prior to the 1940s, research was mainly concerned with predicting performance on a test. With the introduction of statistical measures, such as the well-known correlation tests and factor analysis, more emphasis was placed on the criteria used to predict the validity of a test.

Difficulty in determining this criterion for assessing the validity of new tests in education and psychology led to concepts such as content validity (Rulon, 1946 in Hathcoat, 2013). This approach to validity aligned with the philosophical principle of *psychometric realism,* which relates to the existence of attributes that characterise the test instrument in the real world

(Hathcoat, 2013). This approach suggests that the attributes to be tested exist in the real world, which places psychometric realism within the positivist paradigm ontologically and an objectivist paradigm epistemologically. The main criticism of the psychometric realism principle is that some attributes may be unobservable, and therefore impossible to test using this paradigm only.

Following the above criticism, validity theory shifted its location from psychometric realism to the *construct,* which resulted in the emergence of the seminal work by Cronbach and Meehl (1955). The constructs described may be unobservable, this being based on theoretical concepts that may not be 'real' testable attributes (Hathcoat, 2013). The underlying paradigm therefore was the constructivist's paradigm, in which the focus of validity was not on the test itself, but on the inferences that can be made from the test scores. This also triggered other theorists to explore the inclusion of test interpretation and use of test scores into the validity framing. Messick (1975; 1989; 1998) strongly supported that the 'inferences' from the test scores are important in the validation process.

Messick (1998) proposed a 'constructivist-realist' view (Hathcoat, 2013 p.4) regarding validity theory, which attempts to integrate and unify all types of validity under the overarching label of construct validity, which refers to the integration of evidence that is linked to the interpretation and meaning of test scores. This view of validity and validation offers a theoretical approach to validity that many disciplines in the social sciences, education and health have used in their endeavours to validate test instruments. In this approach, a firm theoretical position must be outlined prior to gathering empirical evidence (Kane, 2011). Messick's unified approach to validity does not discard the empirical layer of evidence gathering, but does allow for the integration of design-related evidence, empirical and measurement related elements, together with the social consequences and outcomes of test scores. This approach also takes into consideration the tension that may exist between the purely positivist approach and the post-positivistic and constructivist approach to validity.

Messick (1998) identifies six main aspects to his constructivist-realist view as content, substantive, structural, generalisability, external and consequential aspects, all of which contribute to the process of validation. Specifically, the content aspect refers to boundaries of the construct that should be investigated, namely:

- substantive aspects: the empirical evidence of the test itself, and how participants respond to the test,
- structural aspects: an acceptable structure of the test that matches what is expected for this construct;
- generalizability: how generalisable the test properties are to other test groups, populations and settings,
- consequential aspect: perhaps the important contribution of this approach is the consequences of the test use, which includes aspects such as test biases and any possible invalidity.

Messick's (1998) approach is not without criticism from contemporary theorists or researchers. Kane (2013) and Borsboom, Cramer, Kievit, Scholten and Franic (2009) argue that Messick's (1998) unified approach to validation lacks a strong theoretical foundation, resulting in it being too broad and open-ended. This argument suggests that Messick's (1998) approach fails to identify at what point the validation process begins and ends, which implies a test can never be truly validated.

Kane (2013) expanded on Messick (1998) to argue that validity is about gathering evidence in an all-encompassing way, and that to deal with the broad conceptualisation of a unified approach to validity, interpretation and uses of the scores should be explicit prior to gathering the evidence. This approach was coined the "argument-based approach" and provides a "framework for evaluating the claims made based on test scores" (Kane, 2013, p. 1). Although this approach has its roots in the earlier theoretical offerings by Cronbach and Meehl and Messick, it more clearly outlines two basic steps in the process: state the claims that are being made explicitly and gather the evidence to evaluate each (Kane, 2011).

Kane's (2013) approach is also known as the interpretive argument approach due to its multiple interpretations and uses. If validity is described as a property of interpretation, then ontologically, this approach does not necessitate the existence of attributes, nor does it support the notion of unobservable attributes (Hathcoat, 2013). Epistemological implications of the argument-based approach indicate that validity is a "tentative judgement based on varying degrees of certainty" (Hathcoat, 2013). In this view, the claim regarding validity is not

necessarily the only ‘truth’, but is a reflection of multiple interpretative judgements about a gathered line of evidence that, according to Hathcoat (2013), may be too broad in its focus.
In contrast to the argument-based approach, Borsboom (2009) presented the instrument-based approach to validity, which emphasises the validity of the test itself over the validity of the process. While the previous approaches focused on validity relating to the use of the test/purpose of the test, this approach specifically focuses on the test itself, with a greater emphasis on the existence of the attributes in the real world in a revisit to psychometric realism. Ontologically, this approach supports the notion that the attributes being tested must exist in the real world where validity is based on the ‘causal relationships’ between the test and specific attribute. In this argument, validation is restricted and does not allow for the consequences of the testing as an aspect of validation. This argument does not support the typical validation practices known to most researchers in audiology.

After a review of the validity theories, an adaptation of Messick’s (1995) unified approach to validity was considered suitable for the validity of the newly developed isiZulu test. Messick (1995) describes a unified validation framework that encompasses the traditional measurement-related validity elements as well as design-related elements to confirm the validity of the test. Central to this idea of validation is the concept of construct validation, which Messick (1995), as cited in Dellinger and Leech (2007), regard as the overarching type of validity. This approach to construct validity is useful, as the validation process allows several levels of evidence to be established relating to measurement and design-related validity, which in turn contributes to the tool’s overall construct validity. Messick describes the aspects of the unified theory on validity in a specific way i.e. content aspect, substantive aspect, structural, generalisability, external and consequential aspects, In this thesis however, the framework describes the measurement related aspects in the traditional way, in that all measurement related elements are part of the overarching construct validity, while still placing an emphasis on the construct.

Figure 3.1 shows this framework as emphasizing measurement-related elements and including reliability testing alongside the four common types of validity testing of criterion, convergent, divergent and content validity. The different levels of evidence gathered for each measurement-related element serve as the empirical layer within the framework, similar to the substantive and structural facet describes in the original Messick framework. In addition, the overall design of the study also allows for design-related elements to be considered, which contribute to the

overall validity of the test scores, including their generalisability and the consequences. In this framework, the evidence gathered for reliability must be viewed collectively with that gathered for the validity for the isiZulu SRT test for it to be used in clinical practice. The collective examination of the scores will contribute to the validation of the isiZulu SRT test. Each of the measurement–related variables will be further reviewed with reference to the speech audiometry test validation, the review starting with the concept of reliability and its types, followed by a review of validity.

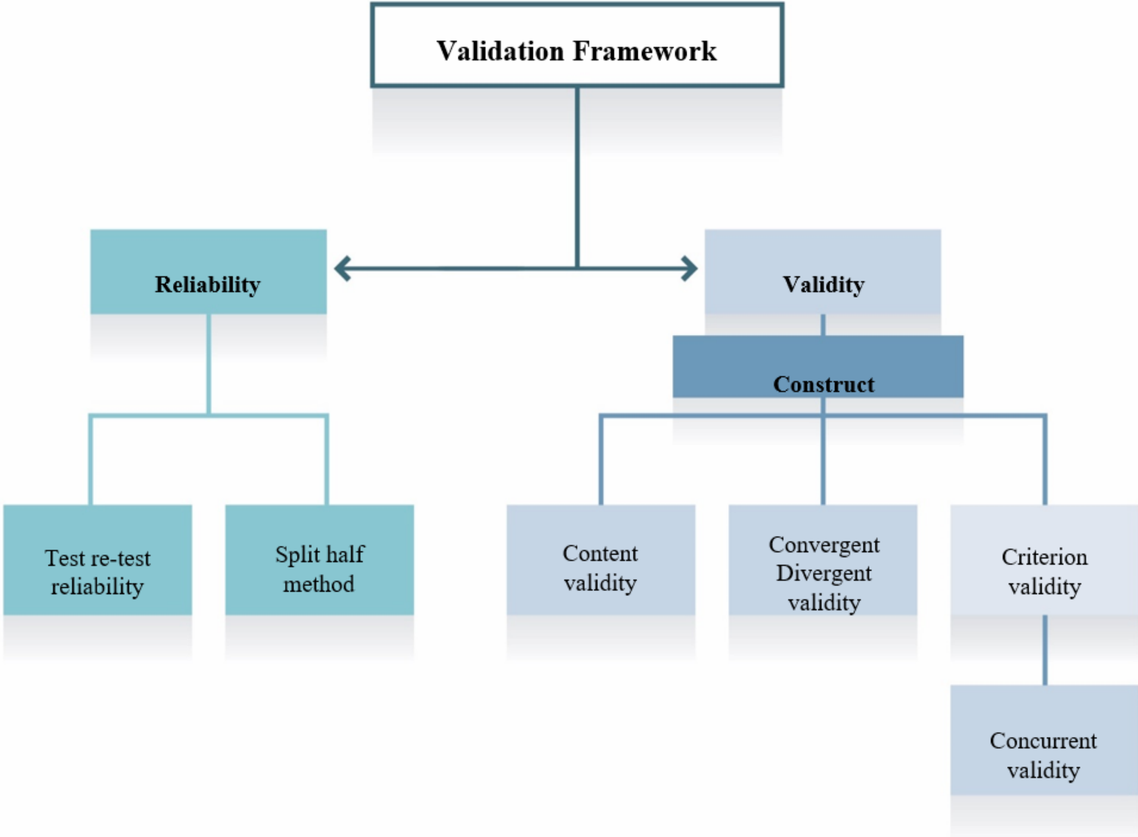

Figure 3.1: Validation framework for this study
Source: adapted from Messick (1995) and Dellinger and Nancy (2007)

3.3 Reliability

The reliability of a test refers to the extent to which its results are repeatable and give the same scores (Joppe, 2000 as cited in Golfshani, 2003; Maxwell & Satake, 2006). Mackie and Demody (1982) and Dillon (1982, as cited in Wang, et al., 2007) stated that four methods have been commonly used to confirm reliability when developing tests for speech recognition: (i) test-retest reliability, (ii) inter–list equivalence or alternate forms, (iii) split-half method, and (iv) intra- and inter-observer reliability, each of which is reviewed.

3.3.1 Test-retest reliability or coefficient of stability

This measure refers to administering the test to the same group of participants on two occasions, and examining the correlation coefficients between the test and retest scores. For hearing tests, the test result remains stable and consistent as long as there is no change to the participants' peripheral and central auditory systems. Therefore, to measure the reliability of a speech audiometry test, such as SRT, the selected participants must have normal peripheral and central auditory systems.

Test-retest reliability evidence exists for some of the speech in quiet and in noise tests developed globally (Hallagren et al., 2006; Neuman et al., 2012). While there have been some reasonably positive reliability outcomes reported for the many speech audiometry tests, various factors could influence their reliability. In 1983, Ostergard provided a review of such factors, and reported that the variance and variability in test scores between tests 1 and 2 in SRT may be influenced first by the number of words in the test list, with a small number of items possibly leading to a more favourable score for SRT. Second, the learning or memory effect is a factor, with the number of words in the lists and the timing of the second test possibly influencing the overall variance. Third, familiarity with the test items and instructions provided may influence the variance, while other factors, such as wide age range in the sample and the test's language, could potentially affect the variance. The question that arises is how small the variance should be for the test to be accepted for clinical practice, with a value of +/- 5dB being accepted in the literature clinically for SRT testing (ANSI, 2004; ASHA, 1988; Caswell, 2013; Hallgren, Larsby & Arlinger, 2006; Neuman et al., 2012).

Notwithstanding many of these factors, the developers of speech-in-noise tests have used test-retest methods to test reliability, and reported good correlation scores. Hallagren et al. (2006) found good test-retest reliability (0.77dB in mean difference was observed and standard deviation was 1.4dB) when 10 participants were tested on the same test by the same tester after one week. A study by Neuman et al. (2012) reported better test-retest reliability ($r = 0.85$, $p<0.05$) for SRT in quiet using sentences when 19 participants were tested and retested four weeks later. By increasing the time, as seen in Neuman et al. (2012), the practice effect may be avoided, with one or more than one-month intervals having traditionally been used to counter the memory effect (Maxwell & Satake, 2006).

More recently, Ng (2016) developed and validated a paediatric word recognition test in Cantonese. The researcher investigated the inter-, intra-rater and test-retest reliability of the 20-item word list on 25 children, the test-retest reliability being found to be good (range of 0.89 - 0.94) using intra-class correlation. The intra-class correlation co-efficient has proven to be a useful method of analysis for reliability studies, as it provides the degree of consistency and absolute agreement among ratings. The findings in Ng (2016) have implications for the present study in term of the method of analysis for a reliability study, as well as the timing between tests 1 and 2. Establishing test-retest reliability was therefore a plausible option for the current study, with other forms of reliability estimates having also been used in the literature for speech audiometry tests, specifically when developing recognition and speech in noise tests, which usually involves testing list equivalence.

3.3.2 Inter-list equivalence or alternate forms of reliability

Inter-list equivalence measures the consistency of a test based on its content, this being achieved by demonstrating that two or more forms of the same test containing equivalent, yet different test items, but can yield the same results. Although the specific test items differ (words, sentences), the rules and criteria for developing these tests are the same (Maxwell & Satake, 2006). Inter-list equivalences for speech audiometry tests imply obtaining equivalence in terms of level of difficulty between the word lists. Equivalence in the difficulty of test lists is established if several lists have been developed to determine speech recognition abilities in quiet or in noise, which is important when participants may require a test to be repeated several times to monitor progress in rehabilitation.

Inter-list equivalence can be achieved by comparing the mean score obtained for each list across participants, i.e. the average threshold obtained across all lists for all participants can be measured and the standard deviations across mean scores calculated. Wang et al. (2007) evaluated the equivalence among 10 lists developed in Mandarin for speech recognition in quiet. In this study, the one Sample Kolmogorov–Smironov statistical test indicated that scores obtained for each list were within the normal distribution curve. Therefore, based on this result, the ANOVA test could be used, as the results indicated a significant difference between the 10 different lists ($p < 0.05$). After further analysis, these authors reported that the scores obtained on list 5 were different from those obtained on their other nine lists, which resulted in list 5 being removed from the next stage of testing. This study is an example of how equivalence in difficulty of test lists can be used to improve the overall reliability of the test measure.

While inter-list equivalence has proven to be a viable option to establish a reliability estimate of the speech tests, it requires several lists (usually more than five) to be available for analysis. The isiZulu SRT test developed by Panday (2006) contains only one list of 26 SRT words, and it is therefore not able to be tested using inter-list equivalence methods of reliability testing. Other forms of reliability testing, therefore, had to be considered, such as internal consistency of the list or the split-half method. Most researchers considered only one type of reliability estimate, while in this study, multiple levels of evidence were gathered to establish the reliability of newly developed tests; the split-half reliability test is another type of reliability estimate considered in this thesis.

3.3.3 The split-half reliability estimate

This type of reliability testing is designed to establish the internal consistency of the test items (Maxwell & Satake, 2006), the method referring to the degree to which scores from half the test items agree with scores obtained from the other half. The original test is split in half and a correlation obtained for the two scores, and as each score is obtained from only half the list, Maxwell and Satake (2006) suggest using a statistical correction factor via the Spearmen-Brown calculation. A further complicating factor is that the entire list contributes to one threshold score, while each item is not correlated. Theunissen et al. (2009) suggest that by correlating the mean threshold values obtained based on a combination of a few words in each list, split-half reliability could also be achieved. As the threshold score for SRT testing is usually obtained by presenting approximately 8–12 words (assuming the initial presentation level was slightly above (10dB) the individual's pure tone average), the 28 words in Panday's (2006) isiZulu SRT test could be split into two groups of 14 words, with a split-half reliability measure possibly be obtained, even with the limitations described above.

3.3.4 Intra-and inter-observer reliability

Intra-and inter-observer reliability have also been used in many studies to establish reliability estimates. Intra-observer reliability refers to the degree of internal consistency of the scores awarded by an individual observer for a prescribed test (Maxwell & Satake, 2006). This differs from inter-observer reliability, which is the proportion of agreement or disagreement between test scores awarded by two or more observers, and is useful when developing the test items for speech audiometry tests. Panday (2006) reported on the use of the Kappa test of agreement to measure the level of agreement among raters when the isiZulu SRT test was administered to normal hearing young adults. The Kappa scores obtained among the raters were greater than

the cut-off score of 0.8, and were therefore considered a good agreement for inter-rater reliability when the isiZulu SRT test was scored for 30 young, normal-hearing adults. This level of reliability was therefore not further evaluated in this study.

From the review of the reliability concept above, the approaches have been largely quantitative, with many of the studies focusing primarily on a single estimate of reliability with small sample sizes. Some of the main factors raised by researchers relate to the learning effect, the period between tests 1 and 2, the wide age range within the sample size and the language of the test itself. These factors have implications for future reliability testing, in addition to other considerations that could improve the overall rigour of the reliability test measurement. Having considered the issue of reliability, the essential validity types and research evidence that exist for validity studies related to speech audiometry are reviewed.

3.4 Validity

Tests that are deemed reliable may not always be valid (Mendel, 2008), and both the reliability and validity of a test should be investigated if it is intended to be used in clinical practice (ASHA, 1988). Reliability and validity measures are interrelated, which has implications for the design and outcomes of the present thesis.

Defining validity has been an ongoing debate in the literature and refers to how well a test or tool measures what it is intended to measure (Maxwell& Satake, 2006; Mendel, 2008). Historically, the concept of validity has been presented as construct, content and criterion validity (Cronbach & Meehl, 1955). In audiology, researchers have focused on some of the measurements related to aspects that are required to confirm the validity of newly developed tests, including criterion, content, or construct validity, with few researchers considering all forms together despite calls to the contrary (McCauley & Swisher, 1984; Mendel, 2008).

3.4.1 Content validity

Content validity refers to how well a particular test adequately represents the various content domains the test was designed to measure (Maxwell & Satake, 2006). The process of establishing content validity is an important part of the initial stages of test development, and typically involves seeking expert opinion and advice to determine if the test items assess what they are required to assess (Maxwell & Satake, 2006). During the construction of the isiZulu SRT test, content domains were considered (Panday, 2007), the criteria for the word list

development were established, and ratings were sought from expert linguists on linguistic criteria, including familiarity, phonetic dissimilarity and tone of the test words. The scores from each of the raters were then assessed for concordance using the Kendall's coefficient of concordance, which identified which of the test words met the stipulated criteria for use in SRT of adult speakers of isiZulu. A W score of 0.8 for familiarity (indicative of very good agreement), 0.7 for phonetic dissimilarity (indicative of good agreement), and 0.43 for tone (considered to be poor agreement) was reported. The words that met the inclusion criteria were used as the test items for the isiZulu SRT test. The content domains considered were also verified with the theoretical guidelines related to criteria for speech reception test development, which provided initial evidence for the content validity of Panday's (2006) isiZulu SRT test.

Other important aspects and characteristics of the isiZulu SRT test word recordings were also investigated, including the naturalness of the recorded words in terms of intonation, tone, clarity and how representative the recording was of the isiZulu dialect. Selecting who should be considered the 'experts' to rate such characteristics of the recording was important for its implications for the current thesis. Other developers of speech tests have used raters/experts to judge the test materials for cultural sensitivity choice of words as it relates to familiarity in the language (Nissen et al., 2005; Nissen et al., 2008), and naturalness of the words to the native language speaker's repertoire (Nilsson, 1994; Wong, 2008 as cited in Theunissen et al., 2009). This not only helps verify the content validity, but also improves the overall face validity of the tests in question and provides some evidence towards being ecologically valid (Theunissen et al., 2009).

Ecological validity in the context of speech audiometry refers to the representativeness of the test material or content to everyday listening conditions (Mackersie, 2002 as cited in Theunissen et al., 2009). Borg, Wilson and Samuelson (1998) state that traditional test procedures, such as SRT in quiet, may not be ideal for evaluating an individual's hearing for speech in real-world listening environments. Similarly, Borg et al. (1998) proposed that to truly assess an individual's hearing for speech, a realistic listening environment must be created and used with appropriate conditions.

In addition, the use of sentence material rather than single words, and the introduction of noise, have been argued as alternative measures to improve the overall ecological validity of speech audiometry testing (Theunissen et al., 2009). While single word tests in quiet may be limited in

terms of assessing how individuals hear in everyday situations, it remains one of the most widely used test methods for speech reception abilities across most clinical settings (Strom, 2003), and is a common practice in most clinical practices internationally. Therefore, given the contextual need for such a test, this thesis focuses on validating a word list for isiZulu in quiet.

3.4.2 Convergent and divergent validity

As part of confirming the theoretical construct, it is important to know if the test returns similar results to other tests that measure the same or a related construct (Schafer, Pogue & Milrany, 2012). This aspect of construct validity is known as convergent validity. To establish convergent validity in speech audiometry, the speech test scores must be correlated to another well-established criterion, such as PTA. As both SRT and PTA tests can be used to predict auditory function for speech within the speech frequencies for each participant, a strong correlation between these scores would argue for convergent validity.

Table 3.1 provides three examples SRT scores correlating with PTA scores in three studies of adult speakers of Telegu, isiZulu and Sesotho. In these studies, the researchers were mainly concerned with correlating the SRT score with those of the PTA for a small group of normal-hearing individuals. The scores for patients with varying degrees and types of hearing loss should also be investigated, a strong correlation within a normative sample being insufficient to argue for the convergent validity of the test. It is also important to determine how the proposed test discriminates from other methods that may not accurately measure the construct, this being known as divergent or discriminant validity. When discriminant validity evidence is gathered, there is usually a poor correlation or a difference in scores between the new test and some other measure that does not reflect the theoretical construct. For this study, a significant difference in the correlation between the gold standard and the new isiZulu SRT, versus between the gold standard and the English SRT test for the same first language isiZulu participants with normal hearing and hearing impairment, will contribute to discriminant validity evidence.

Table 3.1: Review of studies that correlated the SRT and PTA for newly developed word lists

Authors	Year	Target group	Findings
Sreedhar et al. (2011)	2006	50 adult Telegu speakers	Paired SRT words correlated with PTA
Panday (2006)	2006	30 adult isiZulu speakers	IsiZulu SRT test Correlated with PTA (r = 0.76)
Khoza-Shangase & Mokoena (2014)	2014	30 young adult Sesotho speakers	SRT (Sesotho) correlation with PTA (r = 0.64).

3.4.3 Criterion validity

Criterion validity refers to how well a test can predict an outcome, the two sub-types being predictive and concurrent validity. Predictive validity refers to the ability of a given test to predict an outcome that is consistent with the theoretical construct that underlies it. It is usually achieved over a period of time (Maxwell & Satake, 2006), as it seeks to determine if previous test scores correlate with those of future criteria, this aspect of the isiZulu SRT test not being considered in the present thesis.

Concurrent validity refers to how well the test estimates present performance, with evidence in speech audiometry having often been established using psychometric functions that describe the percentage correct identification of a word as a function of its presentation intensity (Kruger & Mazor, 1987). These functions are usually represented by two key parameters: threshold correct level and slope. Threshold correct level is typically defined as the level at which the word is correctly recognised 50% of the time, which determines the location of the word's psychometric function. The correct level for a word should be adjusted in any final SRT test recording so that the threshold level for that word matches the threshold level expected in normally hearing subjects in the target population. Slope is defined as the maximum rate at which performance changes, typically both at the 50% correct level, and between the 20% and 80% correct levels (Liu & Feng Shi, 2013; McPherson & Akeroyd, 2014; Shi & Zaki, 2014). The "slope of a word function expresses the relation between the change in correct recognition performance (Ay) and the change in the presentation level of the signal (Ax) that is expressed as Ay/Ax (%/dB)" (Wilson & Carter, 2001, p7).

The slope needs to be steep, as this provides for more precise estimations of SRT (Hudgins et al., 1947), and greater sensitivity to changes in a listener's ability to understand speech. To achieve a steeply sloping psychometric function, the word must be familiar to the listener and homogenous in audibility (the ease at which the words are understood when spoken at a

constant level of intensity). Several studies have evaluated the psychometric function for selected words in many languages, offering differing models to achieve this measure. Table 3.2 summarises the results of a selection of such studies.

Table 3.2: Studies using psychometric function slope to evaluate and select speech audiometry stimuli in quiet in adults with normal hearing (unless stated otherwise)

Authors	Year	Institute	Test material	Mean slope at 50% correct
Harris et al. (2003)	*2003*	*Brigham Young University*	Bisyllabic Korean SRT	10.35%/dB for male talker and 9.0%/dB for female talker
Harris et al. (2004)	*2004*	*Brigham Young University*	Bisyllabic Polish SRT	10.1%/dB for male talker and 9.8%/dB for female talker
Nissen et al. (2005)	*2005*	*Brigham Young University*	Trisyllabic Mandarin SRT	11.3%/dB% for male talker and 12.1%/dB for female talker
Harris et al. (2007)	*2007*	*Brigham Young University*	Bisyllabic Russian SRT	12%/dB for male talker and 9.9%/dB for female talker
Nissen et al. (2008)	*2008*	*Brigham Young University*	Taiwanese Mandarin Word recognition Testing	6.8%/dB for male talker and 8.3%/dB for female talker
Nissen et al. (2011)	*2011*	*Brigham Young University & University of Hong Kong*	Cantonese Bisyllabic	14.5%/dB 6.8%/dB for male talker and 14.9%/dB for female talker
Wang et al. (2007)	*2007*	*Beijing Institute of Otolaryngology*	Bisyllabic Mandarin word recognition test	2.7%/dB for the hearing-impaired groups and 5.6%/dB for the normal-hearing groups.
Hon Mensah et al. (2023)	*2023*	Ghanaian	Trisyllabic Fante SRT test	7%/dB

Nissen et al. (2005) proposed a mathematical model using a modified logistic regression equation as one method to determine the steepness of the psychometric function slope for selected trisyllabic Mandarin words. In that study, 20 adult native language speakers of standard Mandarin dialect were selected to participate if they presented with normal pure tone and immittance test results. From an initial word pool of 138 trisyllabic words, 90 were selected for high quality, digital recording and evaluation. The psychometric curves for the 90 words were found to be highly variable, but closer inspection revealed that 24 trisyllabic words had steep psychometric functions, the mean slopes of these functions being 11.3%/dB and 9.4%/dB for male and female talkers respectively. As there was still variability in the slopes of the 24 words, the authors suggested a digital adjustment so that the 50% threshold of each word was equal to the mean PTA of the subjects (3.0 dB HL) (Nissen et al., 2005). The mean slope was then recorded to be 11.3%/dB and 12.1%/dB for male and female talkers respectively, with the

digital adjustment. While Nissen et al. (2005) were able to recommend 24 trisyllabic words with fairly steep psychometric functions for SRT in standard Mandarin, they suggested that additional research is needed to compare the SRT test values to the PTA for normal-hearing and hearing-impaired groups. In addition, further studies examining the homogeneity of audibility through psychometric functions should be conducted on the hearing-impaired groups, thereby contributing to concurrent validity data.

Nissen et al. (2008) used the same methodological approach to develop and record high quality digital recordings of bisyllabic Taiwanese Mandarin words, with word recognition testing being the focus. The slope @50% was recorded as 8.3%/dB and 6.8%/dB for female and male speakers respectively (Nissen et al., 2008) when 20 native language speakers of Taiwanese Mandarin with normal hearing were evaluated. These authors report that the slopes found in their study were higher than those in many of the recent English word recognition studies. The difference in the structure of the language between Taiwanese Mandarin and English was considered as a possible reason for steeper slopes in this study (Nissen et al., 2008). In addition, the Nissen et al. (2008) study utilised trisyllabic words, thereby providing the listeners with more acoustic information. These findings, and the postulated reasons provided for the differences in the slopes of the psychometric curves, have implications for the present study. Furthermore, Nissen et al. (2008) recommended that the psychometric curves should be evaluated for varying degrees and type of hearing impairment, and that further test re-test reliability of the developed Mandarin test should be conducted. These recommendations are important for the present investigation, which aims to investigate both reliability and validity parameters systematically for the newly developed isiZulu SRT test.

Similarly, Nissen et al. (2011) digitally recorded and evaluated Cantonese words for speech recognition testing, with trisyllabic words being found to have steep psychometric slopes at 50%, these being reported as 14.5%/dB and 14.9%/dB for male and female talkers respectively. The authors report that although a set of psychometrically equivalent Cantonese words was developed and validated for SRT testing in this normal-hearing population, concern still exists about whether or not this word list would be psychometrically equivalent when hearing-impaired listeners are tested or when used with children. The authors recommended the need to determine the validity of these materials for varying degrees and types of hearing loss and across the age span, which has implications for the present study.

3.4.4 Previous studies investigating the validity of the isiZulu SRT test

Panday et al. (2009) measured the psychometric functions of 58 bisyllabic isiZulu words in search of a homogenous and suitable list of words to measure SRT in that language from those considered in Panday et al. (2007). In total, 131 common bisyllabic Zulu words were identified by two Zulu-speaking language interpreters and two tertiary level educators, with 82% of these words being described as bisyllabic verbs. Using a 3-point Likert-like scale, five linguistic experts then rated (with high agreement) 58 of the bisyllabic verbs as being familiar, phonetically dissimilar and low tone, the words being considered potentially suitable for use in the development of SRT tests in isiZulu.

To evaluate the psychometric functions of the 58 isiZulu words reported above, 30 normal-hearing, adult first language isiZulu speakers with a mean age of 21 years were tested. The isiZulu words were digitally recorded on CD by a male speaker who used the same central dialect of isiZulu as the participants in the study. Logistic regression methods (Nissen et al., 2005, 2008, 2011) were to determine the slope of the psychometric functions of each of the 58 bisyllabic isiZulu words tested. Of these, 28 words showed sufficiently steep psychometric functions for use in SRT testing, with mean slopes at 50% correct, ranging from 1.93 to 15.43%/dB HL (M ± 1SD = 5.99% ± 2.04%/dB), mean slopes at 20% to 80% correct ranging from 1.23 to 9.87%/dB HL (M = 3.83%/dB), and threshold values ranging from −4.30 to 23.60 dB (M = 9.66 dB). The slopes at 50% and the mean slope in this study were found to be more gradual than those reported for other speech audiometry tests (Table 3.2). It was speculated that the structure of the isiZulu language, particularly its tonal properties, the absence of equally stressed syllables and its inclusion of prominent click sounds within its consonant system influenced the threshold correct levels and slopes of the psychometric functions.

Closer inspection of the isiZulu SRT test data, and subsequent analysis using the Mann-Whitney U test, found that when compared to the 53 words without click sounds, the five bisyllabic isiZulu words with click sounds showed similar slopes at 50% correct perception (Z = −1.57, p=0.12) but different thresholds (Z = −2.12, *p*<0.05, with the words containing a click sound having the lower threshold). The better thresholds (Z = −2.12, *p* <0.05) but similar slopes at 50% correct perception seen in the five words with click sounds (versus the 53 words without) suggesting the click sounds could have contributed to lower thresholds of detection but not to higher rates of correct word repetition as stimulus levels were increased. Such an interpretation

should be made with caution, given the differences in sample sizes for words with and without click sounds.

Further evidence for the concurrent validity of the isiZulu SRT test will be determined in the present thesis by examining the psychometric functions (performance-intensity curves) for each isiZulu SRT test word for adult speakers of isiZulu with hearing-impairment. There is a paucity of literature regarding the application of the developed SRT test in quiet for hearing-impaired populations. One such study by Wang et al. (2007) reported a mean slope for the psychometric function of Mandarin words for normal-hearing participants to be 5.6% per dB in comparison to a slope of 2.7% per dB for participants with a sensory-neural hearing loss. This demonstrated previously reported conclusions that performance intensity functions vary with hearing loss (Jerger, 2006; McArdle & Wilson, 2006).

Each of the measurement-related validity elements described in this review are important when considering the validation of a test, with construct validity being central to all these measurement-related elements. Collecting evidence to confirm the theoretical construct of speech perception is important, specifically for speech audiometry tests. Often, with speech perception and speech audiometry being so complex a phenomena, this aspect is not clearly articulated in the literature. Construct validity has been defined as the "degree to which a test or an instrument measures the theoretical concept" under investigation (Maxwell & Satake, 2006 p.115). Construct validity commonly refers to the degree to which a test truly assesses what it supposed to assess. Specifically for speech audiometry tests, establishing construct validity is achieved if the test adequately measures the construct of speech reception. Construct validity is an overarching type of validity that encompasses all other validity types, i.e. content, convergent, divergent and criterion validity (Messick, 1995). This has implications for this study and the way in which evidence was gathered, in that if each of the measurement related elements contributes evidence to the overall construct, the gathering of evidence and its interpretation should be viewed collectively. This concept of collective examination of each measurement-related element is best explored through this integrated framework. Therefore, for this study, to determine the degree of construct validity, the evidence collected for each measurement-related validity was collectively examined.

3.5 Research paradigm and underlying philosophy

The collection of evidence for reliability and validity of the isiZulu test should be carried out in a systematic way, utilising an integrated validation framework. The preceding section provided an overview of the historical and theoretical framework of validity research, based on which the unified approach to validation Messick's proposed was most suitable, with construct validation being at the centre of the framework. Evidence is gathered for all other measurement-related variables to argue for the overall validation of the construct. Therefore, at one level, this study may be considered within a positivistic paradigm, which quantitatively measures the reliability and validity of this test. The test scores are used specifically to indicate correlations that suggest strong relationships and differences. However, given the underlying theoretical framework for this study, a more collective approach to validation is suggested, with a constructive and unified approach appearing to be more suitable. Individual scores for each aspect of reliability and validity were first analysed and then collectively interpreted to confirm the overall validity of the test.

Ontologically, the nature or reality is known because the test scores may be observable, which links well with the positivistic paradigm. However, as the researcher remains objective and observes the participant's performance on the test for much of the study, in view of those involved and various factors that could impact their behaviour, the epistemology shifts, the outcome of the test performance or the validity of the test possibly being influenced by broader factors. Construct realism and the unified view of validity is thus required.

Methodologically, the research followed the rigorous steps required of the positivistic paradigm following an observational, analytical repeated measures design with quantitative methods of analysis. However, positioning or locating this study within a paradigm is not straightforward, as at one level, quantitative data is used to analyse and meet the aims and objectives of this study, but in the analysis process and interpretation of the test scores following the framework suggested for this study, an integrated and holistic view of validity and reliability may shift the study paradigm.

3.5.1 Overview of the Methods

The preceding sections provided the theoretical framework and main measurement-related variables. These are accommodated within the study methodology, with the research paradigm and philosophy as well as the overall research design of the whole thesis being discussed in this

section. In view of this thesis being submitted as a hybrid submission, the methods followed for each of the three studies is presented in the following chapters as part of the publications, However, the relationship between the objectives and methods for each study is presented in Table 3.3 to enable an overview of the different components before the three publications are presented, with the methods for each being detailed therein.

Table 3.3 Overview of the Methods for each of the aims and objectives of the study

No.	Objectives	Characteristics
1.1	**To determine the test-retest reliability of the isiZulu SRT test [External consistency] Part 1**	
	Design	Repeated measures Observational Design
	Participants	• First Language isiZulu Speakers • Normal hearing (Hearing thresholds < 25dBHL • Permanent residents of KZN
	Sampling and recruitment	Convenience sampling *Recruitment*: posters and flyers
	Sample size	23: 4 males, 19 females. Age 21 - 58 yrs, Median age 36 yrs
	Instruments/tools	28 isiZulu SRT words
	Data analysis	• Descriptive statistics • Intraclass correlation (ICC) (2,1) (95% confidence intervals) • Bland Altman Calculation of mean difference between measures ($\bar{d}$), (95% CI for $\bar{d}$,)
1.2	**To determine the (spilt-half reliability) of the isiZulu SRT test [Internal consistency] Part two**	
	Design	Repeated measures Observational Design
	Participants	• First Language isiZulu speakers • Normal hearing (Hearing thresholds < 25dBHL) • Permanent resident in KZN • Age 18 - 60 yrs
	Sampling and Recruitment	Convenience sampling *Recruitment*: posters and flyers
	Sample size	21: 5 males, 16 females, Age 17- 51 yrs, Median 35 yrs
	Instruments/tools	28 isZulu SRT test words and recording
	Data analysis	• Descriptive statistics • ICC (2,1) (95% confidence intervals) • Bland Altman Calculation of mean difference between measures ($\bar{d}$), (95% CI for $\bar{d}$,
2.1.1	**To determine to what extent the SRT test recording represents the natural dialect, intonation and tone of the isiZulu language spoken in KZN**	
	Construct validity	Content validity
	Design	Single observation, cross-sectional design
	Participants	• First language isiZulu speakers. • 18 - 60 yrs • Reside in KZN • Uses dialect spoken in the Durban – Pietermaritzburg region. • Represent various occupational classes & educational backgrounds
	Sampling and recruitment	Purposive Sampling
	Sample size	11: 3 males 8 female, 18 - 60 yrs. Mean age 36 yrs
	Data analysis	Descriptive Stats, Friedman ANOVA p<0.05
2.2.1	**To establish normative data for the isiZulu SRT test**	
	Construct Validity	Convergent, Divergent and Concurrent Validity

	Design	Analytical observational design
	Participants	• First language isiZulu speakers • Normal Hearing thresholds < 25dBHL • eThekwini and surrounding areas
	Sampling and Recruitment	Consecutive sampling. ***Recruitment***: Advertisements in posters and flyers
	Sample size	100 participants; 18 - 60 yrs; Mean age 37 yrs
	Instruments/tools	isiZulu SRT test words and recording
	Data analysis	• Descriptive statistics
2.2.2	**To determine any correlation between the isiZulu SRT score and PTA for hearing (Part 1) and hearing-impaired (Part two) adults (18–60 years).**	
2.2.3	**To determine any correlation between the isiZulu SRT score and PTA for hearing and hearing-impaired adults (18–60 years).**	
2.3.1	**To determine if there is a difference among the isiZulu SRT, English SRT scores and PTA for hearing and hearing impaired adults.**	
2.4.1	**To determine the percentage correct identification for the isiZulu SRT words at specified intensity levels for participants with hearing impairment.**	
2.4.2	**To determine if there is a difference in performance-intensity functions of isiZulu SRT words between hearing and hearing impaired adult groups.**	
	Construct Validity	Convergent, divergent and Concurrent Validity
	Design	Analytical repeated measures design
	Participants	• First language isiZulu speakers • Permanent resident of KZN • Age 18 - 60 yrs
		Part 1: Normal Hearing (hearing thresholds < 25dBHL)
		Part 2: Confirmed hearing loss (mild, moderate, moderately severe and severe)
	Sampling and Recruitment	Part 1: Consecutive sampling *Recruitment*: Advertisements in posters and flyers
		Part Two: Purposive sampling *Recruitment*: Clinical databases of two Audiology Departments in two provincial hospitals and a university clinic in eThekwini, KZN
	Sample size	Part 1: 100 participants. Age 18 - 60 yrs. Mean age 37 yrs
		Part 2: 76 participants
	Instruments/ tools	isiZulu SRT test (zSRT) English SRT test (eSRT) Words were recorded onto CD
	Data analysis	Descriptive statistics Inferential statistics Intraclass correlation (ICC) (2,1), Mixed model analysis p< 0,05 Logistic regression analysis

3.6 Ethical considerations

Full ethical clearance was initially obtained from the University of Cape Town (HREC 652/2012), with subsequent extensions being granted for the entire data-collection phase and during the write-up phase of the study. All procedures conformed to the Declaration of Helsinki principles.

- Gatekeeper permission: Permission was obtained from the KZN Department of Health to access patients from two local hospitals.
- Informed consent: Before any data was collected, those isiZulu speakers who had agreed to participate were provided with an information letter and asked to sign informed consent. At the start of the sessions, they were asked to confirm their voluntary participation and informed about the study and their right to withdraw without any consequences.
- Voluntary participation: potential participants were informed that participation was voluntary, that they would not be paid for their contribution, and that they could stop at any time. They were also informed that they would be provided with money to pay for transport to the testing venue and should they so wish to withdraw, no services would be withheld from them should they decide to withdraw from the study at any stage.
- Anonymity and confidentiality: participants were assured that while some personal details were recorded, including their names for verification purposes, no individual data would be analysed and presented, this being done using summary statistics. They were also informed that at no time would their identities be revealed, and the unique code was allocated to each persons' records to ensure their anonymity and confidentiality.
- Beneficence and Non-malfeasance: Participants benefited from obtaining a free hearing test to be able to be included in the study, and where necessary, were referred for the required medical or audiological follow-up at the local public sector clinic or hospitals, or the UKZN Audiology clinic. Although the study posed minimal risk, participants were briefed on all audiological tests to be administered and the overall benefit of their contribution to the development of an isiZulu test.
- Translation services: While no interpreter was formally employed for the study, this role was filled by a trained research assistant, a first language isiZulu speaker, and a full-time registered audiologist. His responsibilities included giving instructions in isiZulu and English, obtaining informed consent and administering tests. He was extensively trained in the research procedures to ensure that all participants were given the same tests. All documents were translated into isiZulu and back-translated for verification.
- Data security and storage: all hard copy forms were identified with the participants' unique codes to ensure anonymity, and kept by the researcher, being taken each day and locked in the researcher's office for later data entry and analysis. The digital records

were stored on the researchers' password protected computer, with only the supervisor and statistician being given access to the raw data for analysis.

3.7 Conclusion

Having reviewed the historical theories on validity and how the field shifted in philosophy and location, the overarching construction validation framework outlined by Messick (1995) was deemed most suitable for the isiZulu SRT test. This framework allows for the integration of reliability and validity evidence into the validation process, and is sufficiently broad to encompass the important design-related elements for the isiZulu SRT test. The collective examination of both reliability and validity was through this framework, which informed the methodology, with the paradigm and philosophy being relevant to this study. As this is a hybrid thesis submission, the methods for the three studies that address the various objectives are presented in Chapter 4, 5 and 6 as part of the respective publications.

CHAPTER 4. RESULTS OF STUDY 1: RELIABILITY STUDY

Introduction

This chapter reports on the findings for Aim 1 of the study, which focused on the reliability of the isiZulu test, and relates to Objectives 1.1 and 1.2, i.e., to determine the test-retest reliability (external consistency) and split-half reliability (internal consistency of the test). This forms part of Study 1, and addresses the reliability component of the proposed framework.

A manuscript titled "The Internal and External Consistency of the Newly Developed Word Speech Reception Threshold Test for isiZulu Speakers with Normal Hearing Sensitivity" was accepted and published by South African Journal of Communication Disorders in February 2018. This chapter includes the publication in its entirety, with minor adjustments or expansion to the methods section and numbering of figures and tables.

Study 1: Reliability Study

The internal and external consistency of a Speech Reception Threshold (SRT) test for isiZulu speakers with normal hearing sensitivity

[#]**Seema Panday**, *__Harsha Kathard__, [#]**Mershen Pillay** & [^]**Wayne J Wilson**

[#]University of KwaZulu-Natal, Discipline of Audiology

*University of Cape Town, Department of Health and Rehabilitation Sciences

[^]The University of Queensland, School of Health and Rehabilitation Sciences

Contact details of Authors

Seema Panday: pandayse@ukzn.ac.za

Harsha Kathard: harsha.kathard@uct.ac.za

Mershen Pillay: PILLAYM1@ukzn.ac.za

Wayne Wilson: w.wilson@uq.edu.au

Corresponding Author Postal Address:

Seema Panday

University of KwaZulu-Natal (Westville campus)

Discipline of Audiology

Private Bag X54001., Durban, South Africa

Contribution of Authors

Authors Panday, Kathard, Pillay and Wilson contributed conceptually to the study. Panday is the lead investigator of this study and the larger study containing this study. Panday contributed to the data collection; Panday and Wilson contributed to writing, structure and layout; and Panday, Kathard, and Wilson contributed to the editorial aspects of this paper.

Summary This manuscript has 5252 words including the abstract and references; it has 22 pages, 2 tables and 1 figure.

Abstract

Purpose: This study investigated the reliability, particularly the internal and external consistency of a new isiZulu SRT test.

Methods: To examine internal consistency, 21 adult isiZulu speakers with normal hearing sensitivity completed SRT testing using the first and second halves of the SRT wordlist in the same test session. To examine external consistency, another 23 adult isiZulu speakers with normal hearing sensitivity completed the SRT test using the whole word list on two occasions four weeks apart. Consistency of SRT test scores in these test conditions was measured using intraclass correlation co-efficient analyses (a measure of the consistency or reproducibility of different observations of the same quantity) and Bland and Altman analyses of agreement (a comparison of measurement error with the expected variation among subjects).

Results: ICC values ranged from 0.69 to 0.79, showing the isiZulu test scores were highly consistent between the test and re-test conditions used in this study. Bland and Altman analyses showed isiZulu speakers with normal hearing sensitivity can be expected to return isiZulu SRT test scores that differ by no more than 7.5 to 8.7 dB HL between original and repeat assessments.

Conclusion: The isiZulu SRT test was reliable, showing high internal and external consistency when used to assess first language speakers of isiZulu with normal hearing sensitivity. These findings warrant continued development of the isiZulu SRT test for eventual clinical use. This development should include validating this test on first language speakers of isiZulu with and without hearing loss.

Key words: Speech Reception Threshold, internal consistency, external consistency, test-retest reliability, split half reliability

Introduction

Since its introduction in the 1950s, speech reception threshold (SRT) testing has maintained its place in the basic audiological test battery (Ramkissoon, Proctor, Lansing & Bilger, 2002). SRT tests typically consist of a list of words played to a listener who must repeat each word as heard. The level of the words is then altered until the listener correctly repeats 50% of a group of words. This level is the listener's SRT, which is used to quantify the listener's hearing level for speech, cross-check their pure tone average threshold from pure tone audiometry testing, and provide diagnostic and prognostic value for medical, surgical and/or rehabilitative management of hearing loss (Gelfand, 2001).

Despite its widely reported clinical value, SRT testing still faces several significant threats to its reliability and validity as a measure of speech reception. Harris et al. (2007) summarised these threats to include the number, type and homogeneity of the words in the test, the accent or dialect of the speaker, and the method and level of stimulus presentation. Panday, Kathard, Pillay and Govender (2007; 2009) provide a description of how these threats may be addressed when developing SRT tests.

Another important threat to the reliability and validity of SRT testing is the listener's familiarity with the test word stimuli (Eggett & Buck, 2005; Lyregaard, 1997; Nissen, Harris, Jennings). Such familiarity in SRT testing is best achieved by using words from the first language (or native language) of the target population (Nissen et al., 2005). In this regard, Craig (1997) warned the use of unfamiliar test words on groups of first and second language speakers of a language would result in the second language speakers being less able or unable to hear subtle sound segments and prosodic nuances of the word stimuli. Similarly, Takayangi, Dirks and Moshfegh (2002) found that second language English speakers both with and without hearing impairment required higher intensity for equal recognition of English language test words. Such findings suggest that words having a higher frequency of occurrence in the target population, i.e., familiar words are more easily recognised by persons in that population than are words with lower frequencies of occurrence. It also suggests that listeners rely on their higher order cognitive resources, such as prosodic, semantic, lexical knowledge of the language and context for them to recognise and understand test word stimuli (Lyregaard, 1997; Medwedsky, 2002).

The majority of SRT tests used in audiology have been developed in the English language using American English spondee words, this being a bisyllabic word pronounced with equal stress on

each syllable (Ballachandra, 2001; Ramkissoon et al., 2002). The use of these SRT tests on non-English (and even non-American English) speaking populations can be inappropriate, both culturally and linguistically (Ramkissoon et al., 2002). Such inappropriate application can adversely affect the interpretation of SRT test results, with non-American English speakers performing poorly on these tests, not because they have a true speech reception deficit, but because they are simply unfamiliar with American English (Aleksandrovsky, McCullough, & Wilson, 1998; Harris, Kim & Egget, 2003; Harris, Nielson, McPherson, Skarzynski & Egget, 2004; Martin & Hart, 1978; Ramkissoon et al., 2002;).

The potential risks of using culturally and linguistically inappropriate SRT tests are particularly relevant in South Africa, with its diverse multicultural and multilingual population that has experienced major recent socio-political change. Such risks have seen calls to develop contextually relevant tests in South Africa, not only for items such as SRT tests in audiology, but for resources across the full spectrum of health and rehabilitation sciences. One such call for audiology and speech pathology was from Pascoe (2011), who defined contextually relevant resources as “any tools (assessments, intervention programmes, guidelines and norms) that are available for speech-language therapists and audiologists to use with a specific population in a specific setting, and that have been developed with that population and setting in mind.” Importantly, Pascoe also acknowledged efforts in South Africa to not only develop test materials in indigenous local languages, but also to encourage collaborations among researchers, clinicians and the local populations themselves to develop knowledge of the process of these developments.

Attempts to address the need for culturally and linguistically appropriate SRT and other speech audiometry tests in South Africa have generally focused on two areas: the potential for using non-South African language speech audiometry tests and developing new South African language speech audiometry tests. On the potential for using non-South African language speech audiometry tests on South African populations, research has suggested that first language speakers of South African English with normal hearing thresholds perform well on non-South African English speech recognition tests at suprathreshold levels, but poorly at threshold levels. This was seen in studies using the National Acoustic Laboratories Arthur Boothroyd (NAL-AB) wordlists in Australian English (Wilson, Jones & Fridjhon, 1998) and the Central Institute of the Deaf Wordlist 22 (CID W22) in American English (Wilson &

Moodley, 2000). While not addressing SRT testing directly, these studies argue against using or adapting non-South African language speech audiometry tests for South African populations. On the development of new South African language speech audiometry tests, recent research has shown significant promise. This includes the ongoing development of the South African spondaic (SAS) wordlists in South African English (Hanekom, Soer, & Pottas, 2015) and a non-standardised Tswana wordlist (Khoza, Ramma, Mophosho, & Moroka, 2008). Both tests are intended for SRT testing with South African populations, with normal hearing thresholds performing better or equivalently compared to equivalent tests in non-South African languages. It also includes the ongoing development of a speech-in-noise test in Afrikaans (Theunissen, Swanepoel & Hanekom, 2009), the Hanekom et al. (2015) and Khoza et al. (2008) studies in particular supporting the need for ongoing efforts to develop culturally and linguistically appropriate SRT tests for South African populations.

While most South Africans are multilingual, the most spoken language is isiZulu, with approximately 23% of South Africa's 53 million people reporting isiZulu as the primary language they speak at home (Census, 2011). IsiZulu is a Nguni language, with at least two of its differences to the Germanic language of English being of relevance to SRT testing. First, isiZulu is a tonal language (Rycroft & Ncgobo, 1979), where variations in pitch influence word meaning (other examples of tonal languages include Mandarin and Cantonese). Second, isiZulu does not have spondee words, the word type most used for SRT testing in English. Instead, isiZulu favours trochee words ,where prominence is placed on the second syllable of a bisyllabic word. This prominence is often achieved by shortening the duration of the second syllable (Cope, 1982), e.g., the low tone verb "hamba" (meaning "to go") would typically be pronounced with a shortened "ba". The authors refer the readers to Van de Merwe and Le roux (2014) for a review of isiZulu lexical tone and syllable structure.

To address the need for culturally and linguistically appropriate SRT tests for South African speakers of isiZulu, Panday et al. (2007, 2009) developed an isiZulu SRT test, while Panday, Kathard, Pillay and Wilson (2018 a,b) have begun to systematically validate this isiZulu SRT test that considers the linguistic structure of isiZulu. The current version of this isiZulu SRT test consists of a CD recording of 28 common, bisyllabic low tone, isiZulu verb imperatives (spoken by a male first language speaker of isiZulu).

To date, the development of the CD recording of the isiZulu SRT test has concentrated on its content validity, i.e., how well the test represents the content domain it is being designed to measure. To begin, two isiZulu speaking language interpreters and two tertiary level educators identified 131 commonly used isiZulu words for possible use in an isiZulu SRT test, with 124 being subsequently being identified as bisyllabic verbs. Five linguists (each holding a master's degree in the linguistics of African languages) then rated 58 of these bisyllabic verbs as being sufficiently familiar, phonetically dissimilar and low in tone to be potentially suitable for use in the development of an SRT tests in isiZulu (Panday et al., 2007). Recorded versions of these 58 bisyllabic words (spoken by an adult male, first language speaker of isiZulu) were then tested for homogeneity of audibility by playing the words at six intensity levels to 30 isiZulu first language speaking adults (aged 18 to 25 years) with normal hearing (Panday et al., 2009).

Homogeneity of audibility was determined by examining the psychometric functions (also known as the performance-intensity functions or curves) of each of the 58-word recordings. These functions illustrate how well a speech sample or test item is correctly identified as a function of intensity level. Twenty-eight of the recorded words met the criterion of having a mean slope at 50% intelligibility within 1 SD of the group mean of 5.98%/dB. Finally, an analysis of the prosodic features and pitch contours of these 28-word recordings showed they conformed to the prosodic pattern apparent within the linguistic structure of isiZulu (Panday et al., 2009). The present study continues the development and validation of Panday et al.'s (2007, 2009) isiZulu SRT test by determining its reliability when applied to first language isiZulu speakers with normal hearing.

Methods

Research design

This study consisted of two parts. Part 1 was designed to assess the internal consistency and part 2 to assess the external consistency of the new isiZulu SRT test, both using an observational, repeated measures design. An observational design was suitable as participants' responses to speech stimuli were observed, recorded and analysed in an objective way (Leedy & Ormrod, 2013). Repeated measures were also considered as each participant was exposed to both tests for parts 1 and 2, and participants served as their own control (Maxwell & Satake, 2006).

Participants

Part 1 of this study conveniently sampled 21 persons (5 males, 16 females; aged 17 to 51 years, median = 35 years, first quartile = 24 years and third quartile = 44 years), and part 2 conveniently sampled 23 persons (4 males, 19 females; aged 21 to 58 years, median = 36 years, first quartile = 24.5 years and third quartile = 43 years) from the adult, first language, isiZulu speaking population of KwaZulu-Natal (KZN) Province, South Africa. A sample size of 25 participants was targeted for reliability testing, based on the recommendation by Fleis (1986) that a minimum of 20 is required to measure the reliability of a quantitative variable and to compute a correlation statistic, such as the intra-class correlation (with the extra five participants per group recruited to account for attrition). Only 21 persons met the criteria for part 1 and 23 for part 2 of the study.

All participants in this study were first language speakers of isiZulu (self-reported) and permanent residents of KZN (as shown on their South African identity documents). They had responded to advertisements in posters and flyers distributed by the researchers in their communities. After three rounds of advertisements, participants were also recruited through family members of enrolled participants and the University of KwaZulu-Natal community. All participants had unremarkable medical and hearing histories (self-reported), hearing thresholds ≤25 dB HL at octave frequencies from 250 Hz to 8000 Hz, and normal middle ear pressure and compliance (ASHA, 1988; Roup et al., 1998).

The isiZulu SRT wordlist

The ongoing development of the isiZulu SRT wordlist used in this study has been reported by Panday (2006) and Panday et al. (2007, 2009, 2018a). Its current recording (Panday et al. in 2018a) consists of 28 bisyllabic, isiZulu, low tone verbs shown to satisfy criteria of linguistic familiarity and homogeneity of audibility for isiZulu-speaking adults. These verbs are (in order as per the current recording): banga, gxeka, cinga, faka, thela, linda, khaba, kheta, thatha, donsa, washa, chela, xola, yonga, yona, veza, wina, khanya, shada, geza, khipa, thola, jeza, qonda, thenga, loya, minya and yeka. Each recorded isiZulu word had been adjusted in level (Panday et al., 2018) to ensure its 50% correct perception score occurs at the mean pure tone average (2.8 dB HL) of the 30 normally hearing isiZulu first language speaking adults (mean age 21.5 years) who participated in Panday et al. (2009).

Procedure

Ethical clearance was granted by the Faculty of Health Sciences Human Research Ethics Committee of the University of Cape Town to conduct the study (No: HREC 652/2012) (Appendix A). Each participant signed the informed consent form (Appendix B) after reading the information document, which provided information about the study (Appendix C).

Prior to completing the SRT assessments, all participants filled out a case history questionnaire (Appendix D) and underwent pure-tone audiometry assessment at octave frequencies from 250 Hz to 8000 Hz using a Grayson-Stradler GSI 61 twin channel, clinical audiometer with TDH-49 telephonic earphones and MX41-AR cushions, and a tympanometric assessment using a GSI Tympstar clinical middle ear analyser. This testing was completed in an isolated Industrial Acoustics Company twin audiometric soundproof booth of double wall construction, meeting ANSI (1977) standards.

All SRT testing was conducted using the isiZulu SRT test CD, a Technics (SLPG390) CD player, and the audiometer and booth described above. This testing was conducted by an isiZulu first language speaking audiologist with six years' clinical experience who had been trained by the present study's first author (Panday). Each SRT test began with the audiologist showing the participant a printed copy of the words (Appendix E) (in a randomised order) contained in the isiZulu SRT test, as it was being presented on that occasion. The audiologist and the participant then read aloud each word with the participant being given the opportunity to clarify any words with which they were unfamiliar (as recommended by ASHA, 1988). The audiologist then placed headphones on the participant and played the isiZulu SRT test instructions from the CD at 30 dB SL.

To measure each participant's SRT, the audiologist followed a modified version of the Chaiklin and Ventry (1964) descending method, cited in Gelfand (2001), for SRT testing. This method was modified by changing the starting level from 25 dB SL relative to a two-frequency, pure-tone average to 10dB SL relative to the three-frequency (0.5, 1 and 2 kHz) pure-tone average. This was done to reduce test time in expectation of the participants having normal pure tone thresholds.

To complete part 1 of this study of investigating the internal consistency of the isiZulu SRT test, the isiZulu SRT word list was divided into two, with the first 14 words forming Test A

(banga, cinga, thela, khaba, thatha, washa, xola, yona, khanya, geza, thola, qonda, thenga, minya) and the last 14 words forming Test B (gxeka, faka, linda, kheta,donsa, chela, yonga, veza, wina, shada, khipa,jeza, loya, yeka). The SRT for each participant was then determined for Test A and Test B, these SRT measures being conducted sequentially for each participant, and for the right and left ear stimulation separately. The order of testing (Test A then Test B or vice-versa) and ears tested (right then left or vice-versa) was counter-balanced among participants. To complete part 2 of this study investigating the external consistency of the isiZulu SRT test, the SRT for each participant was determined using all 28 words of the isiZulu SRT test on two occasions, four weeks apart, these SRT values being determined for right ear stimulation only.

Data analysis

Descriptive statistics were calculated for each participant's SRT score for each variation of the SRT assessment. These scores were confirmed as meeting parametric assumptions by inspecting their histograms, box-and-whisker plots and Q-Q plots. For both parts 1 and 2, the consistency of the SRT scores obtained was assessed using two methods. The first method was intraclass correlation coefficients (ICCs) using a two-way random model (2, 1) and the 95% confidence intervals (CI) for these ICCs. The second was the Bland and Altman method for assessing agreement, which included calculating the mean difference between measures ($\bar{d}$), the 95% CI for $\bar{d}$, the standard deviation of the differences (SD_{diff}), the 95% limits of agreement, and a reliability coefficient.

The measures of intra-class correlation were completed according to Bartlett and Frost (2008), Rankin and Stokes (1998) and Shrout and Fleiss (1979). The strength of reliability indicated by the ICC values was determined using the general (although arbitrary) guidelines reported by Landis and Koch (1977) of values <0 indicating poor agreement, 0.01–0.20 indicating slight agreement, 0.21–0.40 indicating fair agreement, 0.41–0.60 indicating moderate agreement, 0.61–0.80 indicating substantial agreement, and 0.81–1.00 indicating almost perfect agreement.

The Bland and Altman method of assessing agreement was conducted according to Altman (1991), Bland (1987) and Bland and Altman (1986). The strength of reliability based on this assessment was determined by considering the argument that the reliability of a measure should reflect the true variability of that measure in the target population (Riddle, Finucane Rothstein & Walker, 1989; Streiner & Norman, 1995). This argument suggests that reliability is relative,

and should reflect how well a measurement can differentiate individuals in the target population. In this regard, reliability (or measurement error) should be contrasted with the expected variation among the subjects being tested (Streiner & Norman, 1995). A brief, narrative review of the literature on SRT testing using words suggested a variability of ±5 dB would be a clinically acceptable variation in measured SRT values within subjects (ANSI, 2004; ASHA, 1988; Caswell, 2013; Hallgren, Larsby, & Arlinger, 2006; Neuman, Baumann, Sick, Euler & Weigerber, 2012). As a result, any differences of ±5 dB in the mean difference ($\bar{d}$) scores in any of the SRT assessments conducted in the present study (in the split-half or test-retest analyses) were considered to indicate that the score was reliable, as the difference ($\bar{d}$) score was no greater than the expected variation within subjects.

Ethics

Unconditional ethical clearance was granted by the Faculty of Health Sciences Human Research Ethics Committee of the University of Cape Town to conduct the study (No: HREC 652/2012).

Results

Table 4 shows the SRT measurements and Table 5 the ICCs (2,1) and Bland and Altman test results for parts 1 and 2 of the study, while Figure 2 shows the Bland and Altman plots for parts 1 and 2 of the study. All ICC (2,1) values were in the range of 0.69–0.79, indicating substantial agreement (Landis & Koch, 1977). The Bland and Altman test results showed no evidence of systematic variability in any SRT difference scores. Slight negative trends in the difference values were noted, although they were not considered to be significant relative to the magnitude of the measurement. The observed difference scores ranging from +5 to -5 dB HL would not be regarded as important in clinical measurement of SRT (ANSI 2004; ASHA, 1988; Caswell, 2013; Hallgren et al., 2006; Neuman et al., 2012). The co-efficient of repeatability ranged from 7.5 to 8.7 for the SRT measures in both parts of the study. These coefficients indicate that the differences between two measurements of an SRT value obtained using any one of the three variations of SRT measurement used in the present study can be expected to differ by no more than 7.5 to 8.7 dB HL on 95% of occasions.

Table 4: SRT measurements for parts 1 (internal consistency) and 2 (external consistency) of the study

	Part 1 (split-half)				Part two (test-retest)	
	Test A (1st half of list)		Test B (2nd half of list)		Test (whole list)	Retest (whole list)
	RE	LE	RE	LE	RE	RE
Mean	9.1	9.5	10.2	10.7	10.7	11.1
SD	5.2	4.2	4.0	2.9	4.6	4.51
Minimum	0	5	5	5	5	5
Maximum	20	15	20	15	20	20

Table 5: ICCs (2,1) and Bland and Altman tests for parts one (internal consistency) and two (external consistency) of the study

Part	ICC coefficient	95% CI	$\bar{d}$	SE of $\bar{d}$	95% CI for $\bar{d}$	SD_{diff}	95% LOA	Reliability coefficient
			(dB HL)					
1 (split-half) – RE	0.76	0.49 to 0.89	-1.19	0.68	-2.61 to 0.23	3.12	-7.31 to 4.93	8.7
1 (split-half) – LE	0.69	0.37 to 0.86	-1.19	0.59	-2.42 to 0.04	2.69	-6.47 to 4.09	7.5
2: (test-retest) - RE	0.79	0.57 to 0.91	-0.43	0.62	-1.72 to 0.86	2.98	-6.28 to 5.41	8.3

Note: $\bar{d}$ is the mean difference; SE of $\bar{d}$ is the standard error of the mean difference; 95% CI for $\bar{d}$ is the 95% confidence interval for the mean difference; SD_{diff} is the standard deviation of the differences; LOA is limits of agreement.

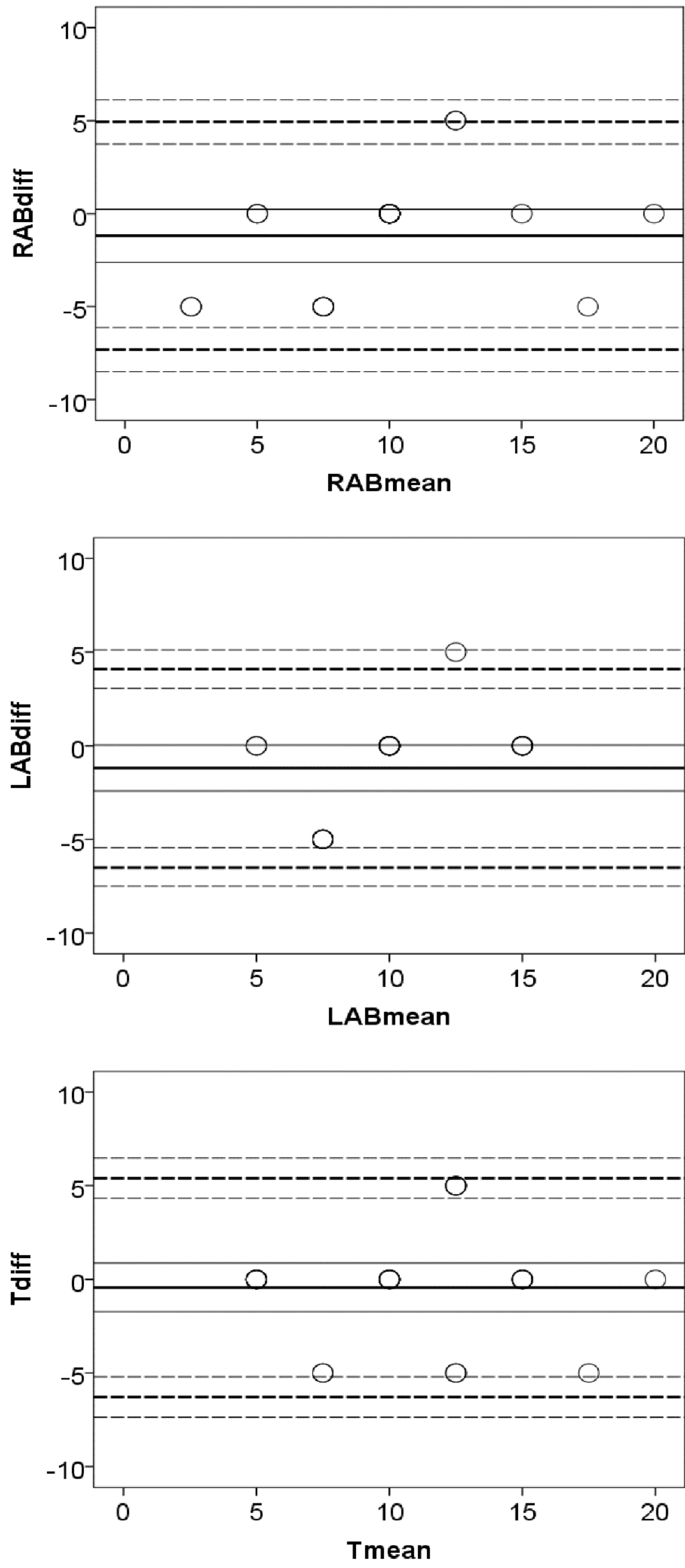

Figure 2: Bland and Altman plots for parts 1 and 2 of the study

Discussion

The isiZulu SRT test was reliable, showing high internal and external consistency when used to assess first language isiZulu speakers with normal hearing sensitivity. The high internal consistency was shown by the split-half analysis results, indicating substantial absolute agreement on ICC (2,1) analyses, and the range of difference ($\bar{d}$) scores being within the ±5 dB variation in SRT scores expected in subjects in a clinical setting on the Bland Altman analyses for both right and left ear stimulation. These results show that both the first and second halves of the isiZulu SRT test wordlist contributed equally to the measurement of SRT in the present study's isiZulu-speaking subjects with normal hearing sensitivity. While not recommended, these results also suggest that the isiZulu SRT test could be used in half-list form (i.e., only half of the words in the list to be used during an assessment) should such a need arise in a clinical setting.

The high external consistency was shown by test-retest analysis results, indicating substantial agreement on ICC (2,1) analysis, and the range of difference ($\bar{d}$) scores on the Bland Altman analyses being within the ±5 dB variation in SRT scores expected in subjects in a clinical setting for the right ear stimulation used in this part of the study. These results show that test scores on the isiZulu SRT test can be expected to remain stable in isiZulu-speaking subjects with normal hearing sensitivity who experience no changes in their hearing over a four-week period. Such a finding should prove useful in clinical settings where serial monitoring of SRTs is needed monthly.

At least four limitations were noted in the present study that should be considered before generalising the results to the wider isiZulu-speaking population. First, the two parts of the study sampled 21 and 23 participants respectively, with both samples being non-random in nature. Second, the method of obtaining the SRT scores was modified from that described by Chaiklin and Ventry (1964) cited in Gelfand (2001), which may not immediately generalise to SRT scores obtained using other methods. Third, the split-half analysis of the isiZulu SRT test was conducted by splitting its wordlist in half by word order (a simple split of the first 12 and the last 12 words in the whole SRT wordlist). Other methods, such as splitting the words by odd and even position, could obtain different results. Finally, the Bland and Altman results were deemed to support the consistency of the SRT test results on the basis that variability of ±5 dB in SRT score would be clinically acceptable. A requirement for less variability would change the interpretation of these Bland and Altman test results.

The high reliability, both for internal and external consistency, of the new isiZulu SRT test showed in first language speakers of isiZulu with normal hearing sensitivity warrants the continued development of this test for eventual clinical use. The 28-isiZulu wordlist recording is not yet available for clinical, use as further research is needed to validate its use on first language speakers of isiZulu with and without hearing loss. As many of the previous studies conducted locally in South Africa focused on the development aspect of speech audiometry tests, the future validity testing of this isiZulu test will contribute to local literature both in terms of methods followed and clinical application of the test for hearing and hearing-impaired individuals.

Declaration of interest

The authors report no conflicts of interest.

Acknowledgements

The authors acknowledge Yusentha Balakrishna from the South African Medical Research Council, Durban, South Africa, and Dr Asad Khan from the School of Health and Rehabilitation Sciences, The University of Queensland, Australia, for assistance with statistical analyses. We also acknowledge Zandile Blose who assisted as a test administrator during the data collection for this study.

Conclusion

This chapter presented the reliability evidence for the newly developed isiZulu SRT test. The internal and external consistency was found to be high, the findings having implications for clinical practice.

CHAPTER 5. RESULTS FOR STUDY 2: CONTENT VALIDITY STUDY

Introduction

This chapter reports on the findings of Study 2, which is the first of the validity studies (content validity study), and addresses Aim 2, specifically Objectives 2.1.1. A paper titled "First Language Raters' Opinions when Validating Word Recordings for a Newly Developed Speech Reception Threshold Test" was published in the South African Journal of Communication Disorders in 2018. The paper is presented as it was published, with minor adjustments or expansion to the methods section and numbering of figures and tables.

First language Raters' Opinions When Validating Word Recordings for a Newly Developed Speech Reception Threshold Test

#Seema Panday, *Harsha Kathard, #Mershen Pillay, ^Wayne Wilson

[#]Seema Panday, *Harsha Kathard, [#]Mershen Pillay & [^]Wayne J Wilson

[#]University of KwaZulu-Natal, Discipline of Audiology

*University of Cape Town, Department of Health and Rehabilitation Sciences

[^]The University of Queensland, School of Health and Rehabilitation Sciences

Received: 08 Mar. 2017; **Accepted:** 05 Oct. 2017; **Published:** 29 Mar. 2018

Abstract

Background: The purpose of this study was to consider the value of adding first language speaker ratings to the process of validating word recordings for use in a new speech reception threshold (SRT) test in audiology. Previous studies identified 28-word recordings as suitable for use in a new SRT test. These word recordings had been shown to satisfy the linguistic criteria of familiarity, phonetic dissimilarity and tone, and the psychometric criterion of homogeneity of audibility.

Objectives: The aim of the study was to consider the value of adding first language speakers' ratings when validating word recordings for a new SRT test.

Method: A single observation, cross-sectional design was used to collect and analyse quantitative data. Eleven first language isiZulu speakers, purposively selected, were asked to rate each of the word recordings for pitch, clarity, naturalness, speech rate and quality on a 5-point Likert scale, with the percentage agreement and Friedman test being used for analysis.

Results: More than 20% of the 11 participants rated the three-word recordings below 'strongly agree' in the category of pitch or tone, and one-word recording below 'strongly agree' in the categories of pitch or tone, clarity or articulation and naturalness or dialect.

Conclusion: The first language speaker ratings proved to be a valuable addition to the process of selecting word recordings for use in a new SRT test. These ratings identified potentially problematic word recordings in the new SRT test that had been missed by the previously and more commonly used linguistic and psychometric selection criteria.

Introduction

Speech reception threshold (SRT) testing is routinely used as part of the basic audiological assessment of hearing (Ramkissoon, Proctor, Lansing & Bilger, 2002). It is typically conducted by presenting a series of words to a listener who must repeat those words as heard. This can be performed using headphones, or in the free field, depending on the age and abilities of the listener. The SRT score is the presentation level (typically reported in decibels hearing level [dB HL]) required for a listener to correctly repeat 50% of a list of presented words. This score is used to quantify the listener's hearing level for speech, and to cross-check the listener's pure-tone audiometry thresholds (the gold standard test for hearing thresholds) (Gelfand, 2001).

Two factors are commonly considered when selecting the words to be included in an SRT test for use in clinical practice: linguistic familiarity and homogeneity of audibility. Linguistic familiarity considers the linguistic properties of words in the language of persons for whom the

SRT test is being developed. It identifies the need to choose words most likely to be familiar to most first language speakers of that language (Kruger & Mazor, 1987). Homogeneity of audibility refers to the "ease at which a word is understood when spoken at a constant level of intensity" (Silman & Silverman, 1991, p. 61). It considers several psychoacoustic properties, including each word's psychometric function and prosodic pattern when spoken (Nissen, Harris, Jennings, Eggert & Buck, 2005).

A word's psychometric function is described as its percentage of correct identification by presentation level (Kruger & Mazor, 1987; Liu & Shi, 2013; MacPherson & Akeroyd, 2014). These functions show that increasing the presentation level of a word increases the probability that a listener will repeat it correctly. The slope of a word's psychometric function should be steep to provide for more precise estimations of SRT (Hudgins, Hawkins, Karlin & Stevens, 1947) and greater sensitivity to changes in a listener's ability to understand speech. A word's prosodic pattern refers to several features, including its length, accent, stress, tone and intonation (Fox, 2000). These features carry different meanings in different languages, and are often determined by measuring the pitch and energy contours of the word (Panday, Kathard, Pillay & Govender, 2009).

Although linguistic criteria and homogeneity of audibility are clearly important in the development of SRT tests in audiology, these factors alone may not be sufficient to ensure that the final SRT test adequately represents all facets of the construct it has been designed to measure (Maxwell & Satake, 2006). In the context of an SRT test, this refers to the need to satisfy not only linguistic and psycho-acoustic criteria, such as those discussed above, but also broader criteria, such as the appropriateness of the word recordings for the communities in which the SRT test will be used. This can be done by including expert speakers of the language from the communities in which the test is recorded. These persons can provide immediate and direct feedback as to the appropriateness of the proposed word recordings being trialled by rating factors such as word intonation, tone and clarity, and representation in terms of dialect. Such ratings could contribute to the final SRT test's content and ecological validity (Theunissen, Swanepoel & Hanekom, 2009), where content validity refers to how well the test words represent the content domain the test was designed to measure, and ecological validity refers to how the test words represent everyday word use in the communities in which the test will be used (Theunissen et al., 2009).

This study reflects on the lessons learnt during the development of a new SRT test for use with speakers of isiZulu in KwaZulu-Natal (KZN), South Africa (Panday, Kathard, Pillay & Govender, 2007, 2009). It considers the value of adding expert speaker ratings of test word recordings where those expert speakers were drawn from the population in which the test is to be used.

Methods

Aim

The aim of the study was to consider the value of adding first language speakers' ratings when validating word recordings for a new SRT test.

Research design

A single observation, cross-sectional design (Maxwell & Satake, 2006) was used to collect quantitative data in this study. The cross-sectional design was suitable as the data was systematically collected at a given point in time, and observational in that participants' responses to speech stimuli were observed, recorded, counted and rated in an objective way (Leedy & Ormrod, 2013).

Participants

The participants were 11 first language, isiZulu speakers (3 men and 8 women, aged 18–60 years with a mean age of 36 years) purposively selected for measuring the suitability of the SRT test word recordings. These participants were permanent residents of KZN Province and were self-reported native speakers of isiZulu with normal hearing thresholds (≤20 dB HL or better at octave frequencies from 125 Hz to 8 kHz), normal middle ear function [within normal limits on acoustic immittance testing] (ASHA, 1990; Roup, Wiley, Safady & Stoppenbach, 1998) and no self-reported history of factors that could affect their hearing ability when listening to the recording of the newly developed word list. They were considered to represent the general population in KZN, with participants coming from the following occupational groups: domestic worker, linguist, student, teacher, technician, librarian, clerk, factory worker and administrator.

Materials and instrumentation

The test word recording

The test word recordings used in this study were the 28 bisyllabic isiZulu low-tone verbs recordings that were selected during previous research for the development of an SRT test for isiZulu speakers in KZN, South Africa (Panday et al., 2007, 2009). This previous study showed

that the 28 word recordings satisfied the criteria of linguistic familiarity and homogeneity of audibility for the target population of isiZulu-speaking adults (Panday et al., 2007, 2009). The criterion of linguistic familiarity had been shown by having two isiZulu-speaking language interpreters and two tertiary-level educators identify 131 commonly used words, with 82% of these subsequently being identified as bisyllabic verbs. Five linguistic experts then rated 58 of these disyllabic verbs as being sufficiently familiar, phonetically dissimilar and low in tone to be potentially suitable for use in the development of an SRT test in isiZulu (Panday et al., 2007).

The criterion of homogeneity of audibility had been shown by recording these 58 words by a male, first language speaker of isiZulu and playing the recordings at six intensity levels to 30 isiZulu, first language speaking adults (aged 18 to 25 years) with hearing within normal limits (Panday et al., 2009). Logistic regression analysis was used to determine the psychometric function of each of the 58-word recordings, with 28-word recordings meeting the criterion of having a mean slope at 50% intelligibility within 1 SD of the group mean of 5.98%/dB. The prosodic features of these 28 word recordings were then analysed and their pitch contours were shown to conform to the prosodic pattern apparent within isiZulu linguistic structure (Panday et al., 2009).

Table 6 shows the 28 disyllabic isiZulu low-tone verbs used in this study and the psychometric properties of each word reported in previous research (Panday et al., 2009). For this study, the levels of each of the 28-word recordings were adjusted (ΔdB) so that the 50% correct perception scores for each word occurred at the mean pure-tone average (2.8 dB HL) of the participants in the previous homogeneity of audibility study. The size of this adjustment for each word is shown by the ΔdB levels in Figure 3 indicates each recorded word's psychometric function before and after this level adjustment.

Table 6: Psychometric properties of the 28 disyllabic isiZulu word recordings used in this study

Word	Slope at 50%	Slope from 20% to 80%	Thresh hold	ΔdBb
qonda	15.43	9.87	4.18	1.38
thola	10.98	7.02	5.27	2.47
washa	9.58	6.13	2.90	0.10
thenga	8.83	5.65	8.13	5.33
jeza	8.05	5.15	9.13	6.33
yona	8.05	5.15	9.13	6.33
xola	8.03	5.14	7.05	4.25
yeka	7.90	5.06	6.52	3.72
chela	7.68	4.91	9.82	7.02
shada	7.60	4.86	6.51	3.71
gxeka	7.50	4.80	3.74	0.94
khaba	7.20	4.61	6.99	4.19
veza	6.83	4.37	14.44	11.64
thela	6.78	4.34	10.62	7.82
yonga	6.53	4.18	7.31	4.51
kheta	6.50	4.16	8.57	5.77
wina	6.50	4.16	13.08	10.28
cinga	6.45	4.13	11.28	8.48
faka	6.43	4.11	1.10	3.63
khipa	6.35	4.06	10.16	7.36
thatha	6.05	3.87	11.11	8.31
linda	6.00	3.84	9.45	6.65
loya	5.93	3.79	12.73	9.93
minya	5.93	3.79	15.45	12.65
donsa	5.85	3.74	9.25	6.45
khanya	5.73	3.66	6.05	3.25
geza	5.70	3.65	8.11	5.31
banga	5.68	3.63	10.54	7.74

a, Mean presentation level (dB HL) required for 50% correct perception; b, change in mean presentation level (dB) required to adjust the 50% correct perception threshold of a word to the mean three-frequency (0.5 kHz, 1 kHz and 2 kHz) pure-tone average (2.8 dB HL) of all participants (all changes required a decrease in level by the indicated amounts); c, words containing click sounds.

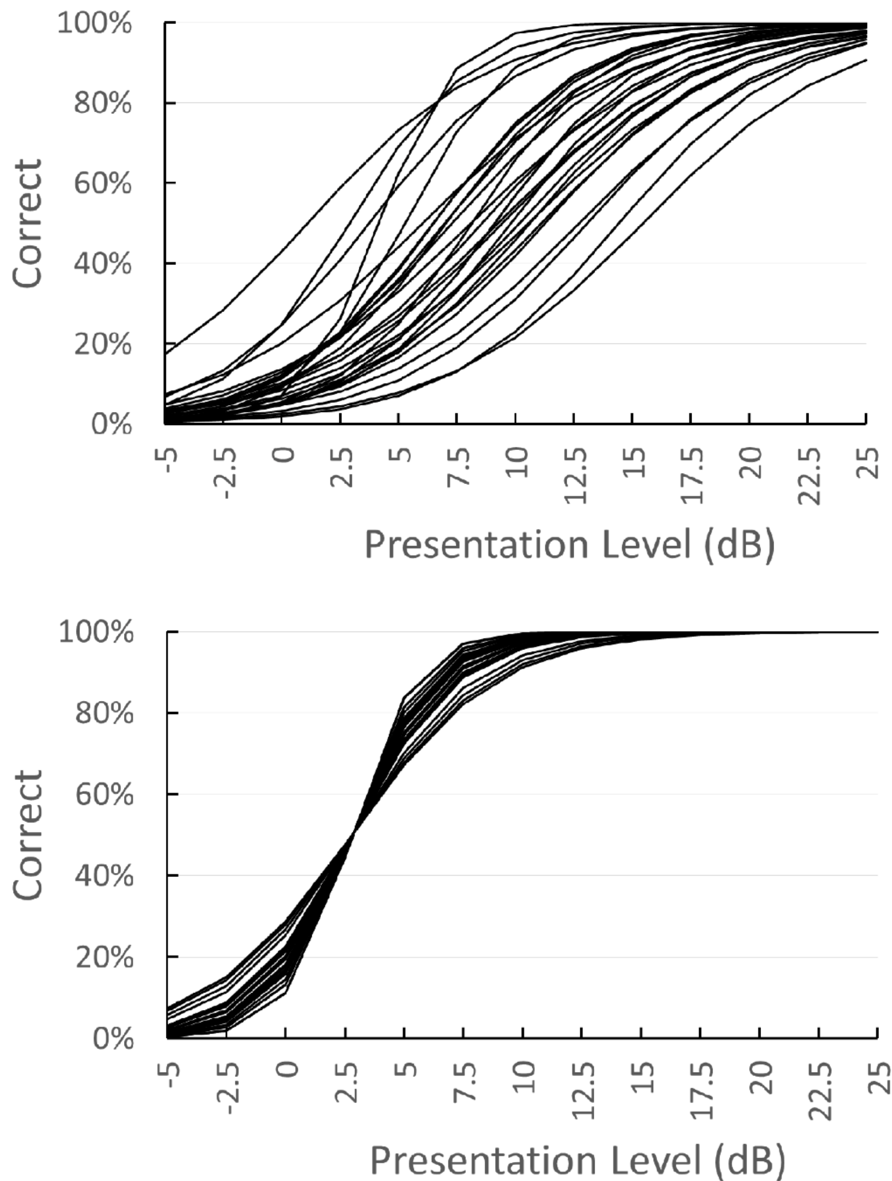

FIGURE 3: Psychometric functions for the 28 recorded isiZulu words unadjusted (a) and adjusted (b) for listener hearing thresholds

The rating scales

The rating scale used in this study was developed based on previous research (Theunissen et al., 2009) (Appendix F). It consisted of five questions asking the participant to rate each word on five criteria: pitch or tone, clarity/articulation, naturalness or dialect, speech rate and quality. Participants responded to each question on a 5-point Likert-like scale,where 1 corresponded to strongly agree, 2 to agree, 3 to neutral, 4 to disagree and 5 to strongly disagree.

Data collection procedure

A pilot study was first conducted on three participants who were not part of the main study. The rating scale and test procedure were piloted for accuracy and ambiguity, with minor editorial and rephrasing of instructions being made, after which the participants for the main study were recruited.

Each first language, isiZulu-speaking participant in the main study was shown the rating scale and instructed on its use. Each participant then listened to each of the 28 recorded isiZulu word recordings. Each word recording was played three times at a comfortable listening level of 60 dB HL before each participant was asked to complete their ratings of that word recording. The word recordings were routed from a Technics (SLPG390) CD player to the external input of a Grason Stadler model 61 audiometer (calibrated in accordance with ANSI S3.6 specifications [ANSI, 2004]) and then to the test ear (chosen by flipping a coin) of each participant through a single TDH-49 headphone with MX41/AR cushions. Prior to testing each participant, the external inputs to the audiometer were calibrated to 0 VU using the 1000 Hz calibration tone on track one of the test CD.

Data analysis

The ratings, obtained from the 11 first language isiZulu-speaking participants about the pitch or tone, clarity/articulation, naturalness or dialect, speech rate and quality of speech for each of the 28 isiZulu word recordings were analysed descriptively and inferentially. Agreement amongst the 11 raters was calculated using percentage, which were exceptionally high, further analysis using kappa statistics was abandoned in favour of the Friedman analysis of variance (ANOVA) analyses (McHugh, 2012). These Friedman ANOVA analyses considered the rater as the independent variable and the ratings in each rating category as the dependent variable. A Friedman ANOVA analysis was conducted for each rating category separately (pitch or tone, clarity/articulation, etc.), with all statistical analyses were carried out using Stata version 14 under the guidance of a statistician from the Medical Research Council of South Africa (Durban).

Ethical consideration

Unconditional ethical clearance was obtained from the University of Cape Town (HREC 652/2012), verbal and written informed consent were obtained from all participants, who were given an information document that outlined the details of the study (Appendix C). Participants

were assured of their anonymity and were free to withdraw from the study at any point, with. confidentiality being maintained throughout the study.

Results

Figure 4 shows ≥80% of the 11 first language isiZulu-speaking participants rated 24 of the 28 isiZulu word recordings as 'strongly agree' in the categories of pitch or tone, clarity/ articulation, naturalness or dialect, speech rate and quality. More than 20% of these 11 participants rated the three-word recordings, 'kheta', 'washa' and 'yonga', below 'strongly agree' in the category of pitch or tone, and one-word recording, 'cinga', below 'strongly agree' in the categories of pitch or tone, clarity or articulation and naturalness or dialect.

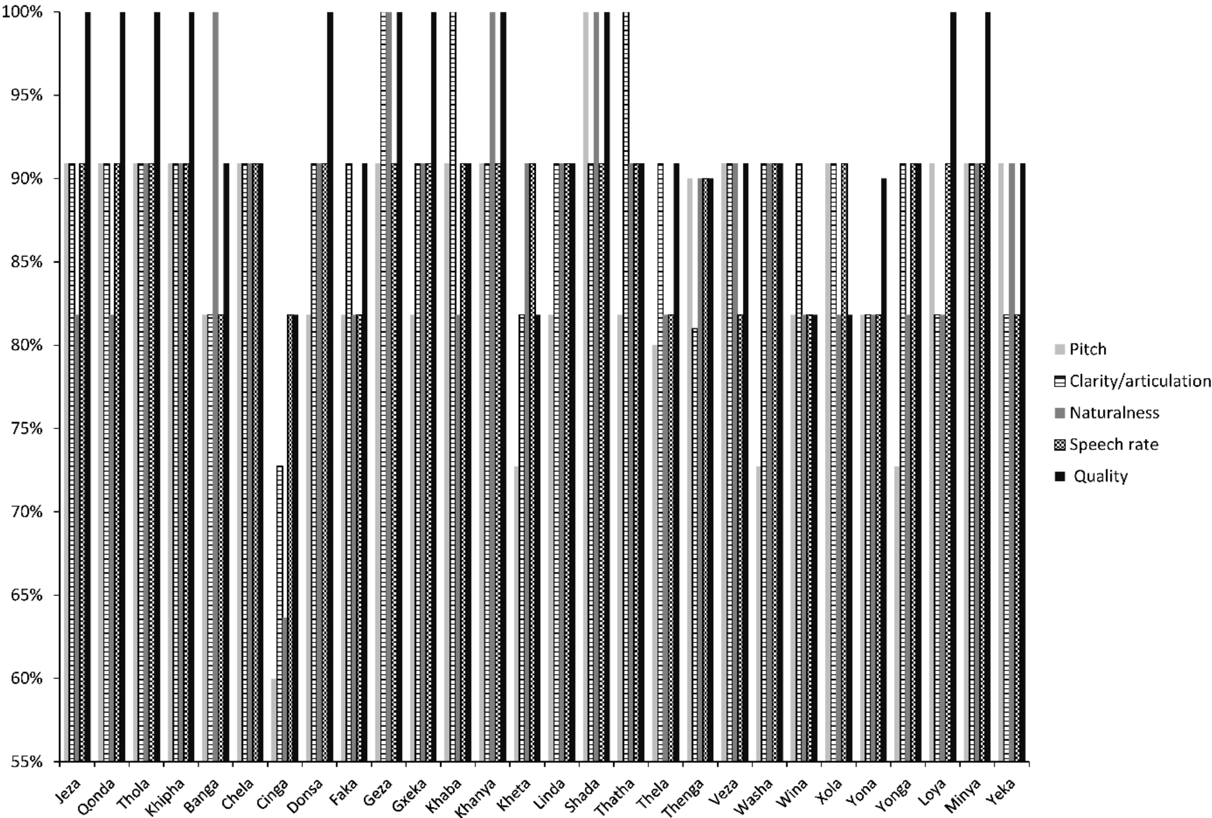

Figure 4: Percentage of raters strongly agreeing in each rating category for each of the 28 recorded isiZulu word

The assessment of rater reliability showed that the chi-square statistic of 18.31 (df = 10) was far greater than the calculated Friedman statistic in each rating category (pitch or tone, clarity/articulation, etc.) This led to the null hypothesis being accepted, that being that there was agreement among the 11 participating raters within each of the five categories at a 95% confidence level.

Table 7 Friedman analysis of variance analyses comparing raters within each rating category

Category	Friedman statistic	*p*
Pitch or tone	5.16	0.9
Clarity/articulation	3.05	0.9
Naturalness or dialect	4.73	0.9
Speech rate	2.07	0.9
Quality	3.44	0.9

Discussion

The first language speaker ratings were a valuable addition to the process of validating word recordings for use in a new SRT test. This was evident in at least four aspects of the rating scale:

(1) by keeping the scale simple, it proved to be easy to develop;

(2) by targeting the population for whom the SRT test was being developed, the participants proved to be easy to recruit;

(3) with points one and two in place, the participants found the scale easy to complete, and

(4) the results obtained from the scale were easily integrated with the linguistic and psychometric results already obtained for the SRT in its development to date, and added evidence for the content validity of this test.

The first language speaker ratings were also a useful addition to the process of validating word recordings for use in a new SRT test. This was most evident in all 28 word recordings having been previously accepted on linguistic and psychometric criteria (Panday et al., 2007, 2009), the first language speakers then identifying potential problems with three word recordings on pitch or tone, and one word recording on pitch or tone, clarity or articulation and naturalness or dialect. A possible cause for these problems could lie in the way the first language isiZulu speaker spoke the words during the original recording. On reviewing these recordings, this speaker was heard to have provided greater emphasis on their second syllable (an iambic meter), which resulted in a slight change in the pitch and tone of these word, particularly for the word 'cinga'. It was noted that this iambic meter (greater emphasis on the second syllable) differs from the spondaic meter (equal emphasis on each syllable) present in the disyllabic word recordings used in English language speech reception tests.

The value added by including first language speaker ratings in the development of the isiZulu SRT test, considered in this study, supports suggestions that commonly used measures of linguistic criteria and homogeneity of audibility alone may not be sufficient to ensure that the

final SRT test adequately represents all facets of the construct it has been designed to measure (Nissen et al., 2005; Maxwell & Satake, 2006). Considering broader criteria, such as the appropriateness of the word recordings, as rated by members of the communities in which the SRT test will be used, can provide immediate and direct feedback as to the appropriateness of the proposed word recordings being trialled. Such ratings will contribute to the content and ecological validity of any final SRT test recording. Pascoe, Rogers and Norman (2013) indicate that there is need for researchers in South Africa to share the methods used to develop and validate contextually relevant tests. The use of community members or first language speakers as experts of the language when selecting word recordings for SRT tests is another method to consider when developing contextually relevant materials in South Africa.

Conclusion

The first language speaker ratings proved to be a valuable addition to the process of selecting word recordings for use in a new SRT test. These ratings identified potentially problematic word recordings that had been missed by the previously and more commonly used linguistic and psychometric selection criteria. This study has implications for the methods used to determine the reliability and validity of new tests for speech audiometry. It supports the use of multiple methods to systematically accumulate evidence for and against the use of new tests, and viewing this evidence in an integrated way to determine the true reliability and validity of new tests in the settings in which they will be used.

Acknowledgements

The authors thank the South African Medical Research Council for biostatistical support, specifically Cathy Connolly, Rebecca Shunmugam and Yusentha Balakrishna.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

S.P., H.K., M.P. and W.W. contributed conceptually to the study. S.P. is the lead investigator of this study and the larger study containing this study. S.P. contributed to the data collection; S.P. and W.W. contributed to the writing, structure and layout; and S.P., H.K., M.P. and W.W. contributed to the editorial aspects of this article.

CHAPTER 6. VALIDITY STUDY 3

Introduction

This chapter reports firstly on the findings of the convergent validity (Objective 2.2.2) and then on the concurrent validity of the newly developed test when participants with hearing and without hearing loss were tested (Objective 2.4.1). It is presented as a paper that was published in the South African Journal of Communication Disorders in November 2020. This chapter also reports on the normative and divergent validity findings as additional sections to this chapter (sections 6.1, 6.2), and summarises the divergent validity (Objective 2.3.1) and normative findings (Objective 2.2.1).

The Validity of an isiZulu Speech Reception Threshold Test for Use with Adult isiZulu-Speakers

[#]Seema Panday, *Harsha Kathard, & [^]Wayne J Wilson

[#]University of KwaZulu-Natal, Discipline of Audiology

*University of Cape Town, Department of Health and Rehabilitation Sciences

[^]The University of Queensland, School of Health and Rehabilitation Sciences

Contact details of authors

Seema Panday: pandayse@ukzn.ac.za

Harsha Kathard: harsha.kathard@uct.ac.za

Wayne Wilson: w.wilson@uq.edu.au

Corresponding author postal address:

Seema Panday

University of KwaZulu-Natal (Westville campus)

Discipline of Audiology

Private Bag X54001, Durban, South Africa

Contribution of authors

Authors Panday, Kathard and Wilson contributed conceptually to the study. Panday is the lead investigator of this study and the larger one containing this study. Panday contributed to the

data collection; Panday and Wilson contributed to writing, structure and layout; and Panday, Kathard and Wilson contributed to the editorial aspects of this paper.

Summary This manuscript has 4 731 words including the abstract and references; it has 25 pages; 4 tables and 1 figure.

Received 31 October 2019; **Accepted:** 16 July 2020; **Published**: 09 November 2020.

Abstract

Background and objectives:

This study continued the development of an isiZulu speech reception threshold (zSRT) test for use with first language, adult speakers of isiZulu. The objective was to determine the validity of the zSRT test for use with adult isiZulu speakers.

Methods: Two participant groups of adult, isiZulu first language speakers with normal hearing ($n_1 = 100$) and adult, isiZulu first language speakers with hearing loss ($n_2 = 76$) were assessed on pure tone audiometry and the zSRT test. Convergent validity was established through agreement of the zSRT scores with pure tone average (PTA) scores, and concurrent validity was assessed by examining the steepness of the psychometric curve for each word in the zSRT test for each type and degree of hearing loss.

Results: Intraclass correlation coefficient analyses showed zSRT scores ranged from substantial to very high agreement with PTA scores for the normal hearing and hearing loss groups (NH – right ear $ICC_{consistency} = 0.78$, left ear ICC = 0.67; HL – right ear $ICC_{consistency} = 0.97$, left ear $ICC_{consistency} = 0.95$). The mean psychometric slope (%/dB) at 50% correct perception for all words in the zSRT test was 4.92%/dB for the mild conductive hearing loss group, 5.26%/dB for the moderate conductive hearing loss group, 2.85%/dB for the moderately severe sensorineural hearing loss group, and 2.47%/dB for the severe sensorineural hearing loss group. These slopes were appropriate for the degree of hearing loss observed in each group.

Conclusion: The isiZulu SRT test showed convergent and concurrent validity for assessing SRT in first language, adult speakers of isiZulu with hearing loss.

Keywords: Speech Reception Threshold, isiZulu, convergent validity, concurrent validity

Introduction

This article reports on the convergent and concurrent validity of an isiZulu speech reception threshold (zSRT) test for use with first language, adult speakers of isiZulu with and without hearing loss. With evidence-based practice now being prioritised for clinical practice, the need for validity evidence of newly developed tests continues to grow among clinicians and researchers alike (Friberg & McNamara, 2010). In the case of speech audiometry, evidence of the validity of its tests has been dominated by reports of the psychometric functions (also called the performance intensity functions) for the word stimuli used in these tests (e.g., Harris et al., 2001, 2003, 2004; Nissen et al., 2005, 2007). While such reports provide some evidence of validity and reliability, systematic assessments of the validity of speech audiometry tests remain rare. This is concerning, as the clinical use of any test in audiology should be based on a rigorous, strict and systematic investigation of its validity and reliability, the lack of such investigation limiting the use of these tests in clinical practice (Ma, McPherson & Ma, 2013).

In general, validity refers to the degree to which a test or measurement tool achieves what it is supposed to measure (Maxwell & Satake, 2006; Mendel, 2008). Historically, validity has been presented as separate subtypes (Cronbach & Meehl, 1955) including construct validity, that being the degree to which a test or instrument measures the theoretical construct under investigation. It is recommended that all subtypes of validity be considered for test validation, with authors such as Messick (1995) offering unified validation frameworks to achieve this goal. Central to this framework's consideration of validity is the placement of construct validity as the overarching type of validity (Messick, 1995, in Delinger & Nancy, 2007). Construct validity is then broken down into content, convergent and divergent, and criterion/concurrent validity. Messick's (1995) framework is useful as it allows several levels of evidence to be gathered to argue for or against the overall validity of any new test. Specifically for this paper, evidence relating to convergent and concurrent validity is the focus.

Speech reception threshold (SRT) testing remains a core component of the audiological test battery (Ramkissoon, Estis & Flagge, 2014). On hearing a list of words that have been pre-recorded or presented by live voice, the listener is asked to repeat each word as heard. By scoring the responses, an examiner can determine the listener's SRT as the level at which they correctly repeated the words 50% of the time. This SRT score can then be used to quantify the listener's speech reception ability, to cross-check other audiometric results, such as the pure

tone average threshold obtained from pure tone audiometry, to confirm or deny sites-of-lesion along the auditory pathway, and to guide some forms of auditory rehabilitation (Gelfand, 2001). While SRT testing has been used in audiology clinics around the world since the 1950s, it continues to face several challenges. Perhaps the largest of these is the need to match the words used in an SRT test to the contextual, linguistic and clinical factors relevant to the target population. Such a challenge is of particular relevance in a country such as South Africa that has 12 official languages: Afrikaans, English, Ndebele, Northern Sotho, Sotho, Swazi, Tsonga, Tswana, Venda, isiXhosa and isiZulu.

At present, the majority of SRT tests have been developed in the English language, particularly in American English (Ballachanda, 2001; Ramkissoon, 2002). Use of these tests in South Africa has some (albeit limited) prima facie value with South African English being spoken as the primary language at home by 8.1% of the population (Statistics South Africa, 2018), making it the sixth most common primary spoken language in South Africa after isiZulu (23%), isiXhosa (16%), and Afrikaans (14%). South African English is also one of several languages commonly used in most urban areas in South Africa (Statistics South Africa, 2018), and is the dominant language spoken in government and media settings. Despite this, it remains clear that the use of SRT tests in English (and the dominant American English) on the South African population is culturally and linguistically inappropriate at best (Ramkissoon, 2002) and invalid at worst (Aleksandrovsky, McCullough & Wilson, 1998; Harris et al., 2003; Harris et al., 2004; Martin & Hart, 1978; Ramkissoon et al., 2002).

In response to the need for valid speech audiometry tests in South Africa, researchers have begun to develop locally relevant tests (Naude, 2018; Hanekom, Soer, & Pottas, 2015; Khoza, Ramma, Mophosho, & Moroka, 2008). This includes the development of an SRT test in the most spoken language in South Africa, isiZulu (Census, 2011) (Panday, Kathard, Pillay & Govender, 2007, 2009; Panday, Kathard, Pillay & Wilson 2018a; Panday, Kathard, Pillay & Wilson, 2018b). The current recording of this isiZulu SRT test consists of 28 common, bisyllabic isiZulu words spoken by a male, first language speaker of isiZulu. These words have been shown to be linguistically familiar and homogenous in audibility in first language, adult speakers of isiZulu (Panday et al, 2007, 2009, 2018a,b), this test having been shown to produce reliable SRTs when used on adults with normal hearing sensitivity (Panday et al, 2018a). If this new isiZulu SRT test is to be used on first language, adult speakers of isiZulu in South Africa, then its validity when applied to individuals with and without hearing loss must be assessed.

The aim of this study was therefore to determine whether the isiZulu SRT test is a valid measure of SRT in first language, adult speakers of isiZulu speakers with and without hearing loss.

Methods

Research Design

This study used an analytical, observational design (Maxwell & Satake, 2006) to collect and analyse quantitative data in two parts, one involving participants with normal hearing and one involving those with hearing loss.

Participants

All participants were first language, adult speakers of isiZulu (self-reported) and permanent residents of KwaZulu-Natal (KZN) Province. For part 1 of the present study, 100 participants with normal hearing aged 18 to 60 years (mean age 37 years) were recruited through advertisements and selected based on consecutive sampling (Maxwell & Satake, 2006) from the eThekwini Municipality and surrounding areas of KZN. Participants had to have unremarkable case history and their hearing status had to be confirmed by audiometric testing. For part two of the study, 76 adults with hearing loss between the ages of 21 to 59 years (mean age 39 years) were purposely sampled from the clinical databases of two Audiology Departments in two provincial hospitals and the Audiology Clinic of a university in the EThekwini Municipality, KZN. The hearing status of each participant was identified from their hospital or clinical database records and confirmed by audiometric assessment at the research facility. The 76 participants were grouped by the degree of hearing loss in their better hearing ear and resulted in four groups: mild conductive hearing loss (26-40 dB, n =15), moderate conductive hearing loss (41–55dB, n =20), moderately-severe sensorineural hearing loss (56–70dB, n =21), and severe sensorineural hearing loss (71–90dB, n = 20).

The isiZulu SRT test

The isiZulu SRT test used in this study has been reported in detail by Panday et al. (2007, 2009, 2018a,b). It consists of 28 bisyllabic isiZulu low-tone verbs that are linguistically familiar and homogenous in audibility for isiZulu-speaking adults. The words were recorded to a compact disk with a calibration tone and test instructions (in isiZulu) for the listeners.

Test administrator

All testing was conducted by an audiologist with six years' clinical experience who had been trained by the present study's first author (Panday), and was first language speaker of isiZulu.

Procedure

All testing was done in an isolated Industrial Acoustics Company twin audiometric soundproof booth of double wall construction meeting ANSI (1977) standards using a Grayson-Stradler GSI 61 twin channel clinical audiometer with TDH-49 Telephonics earphones and MX41-AR cushions, a Technics (SLPG390) compact disk player, and a GSI Tympstar clinical middle ear analyser. Testing occurred in two sessions. Participants in part 1 of the present study only participated in testing session one, while those in part 2 participated in testing sessions 1 and 2.

Session 1

In the first session, all participants completed pure-tone audiometry, tympanometry and SRT testing using the zSRT test. Pure tone audiometry was conducted at octave frequencies from 250 Hz to 8000 Hz using the modified Hughson-Westlake threshold technique (Carhart & Jerger, 1959). The PTA (the average hearing threshold for 0.5, 1 and 2KHz) and the zSRT were calculated and recorded.

For the isiZulu SRT testing, each participant listened to the words being read aloud by the tester and was given the opportunity to clarify any unfamiliar words (as recommended by ASHA, 1988). A modified version of the Chaiklin and Ventry (1964) descending method (as cited in Gelfand, 2001) was used to conduct the SRT testing proper. This method was modified by changing the starting level from 25 dB SL relative to a two-frequency pure-tone average to 10dB SL relative to the three-frequency (0.5, 1 and 2 kHz) pure-tone average. If the participant was unable to repeat the initial word played, then the SRT test sequence was restarted by presenting a new word at the initial presentation level plus 10 dB. Once the participant was able to repeat a word, then a further two words were presented at the same presentation level. If all three words were not correctly repeated, the audiologist increased the presentation level by 5 dB and played the next three words. Once all three words were correctly repeated, the audiologist reduced the presentation level by 5 dB and played the next three words.

This was repeated until the participant incorrectly repeated any one of the words in a three-word block, at which point the audiologist played another three words at the same level to get a score out of six. If the participant correctly repeated three out of the six words, then the presentation level of those words was deemed to be the participant's SRT and the SRT testing was stopped. If the participant correctly repeated fewer than three out of the six words, then the presentation level was increased by 5 dB and another set of six words were presented. If the participant correctly repeated more than three out of the six words, then the presentation level was decreased by 5 dB and another set of six words were presented. This

sequence of increasing or decreasing the presentation level was repeated until the participant correctly repeated three out of a six-word block, the presentation level of which was deemed to be the participant's SRT, at which the testing was stopped. In sum, a participant's SRT was deemed to be the lowest presentation level at which they correctly repeated three out of a block of six words.

Session 2

In the second sessions, only participants with hearing loss completed word recognition testing using the zSRT test words that was played at varying presentation levels to obtain psychometric functions for each word in the test. Each participant was instructed in isiZulu by the test administrator that a series of words would be played through the headphones to their better hearing ear (or to the right ear if the hearing was symmetrical), their task was being to repeat each word as they heard it.

All 28 words on the isiZulu SRT test were first played at 20 dB SL to familiarise each participant with the words, and then in a randomised order at 35 to 70 dB HL (dial setting) in 5 dB steps for participants with mild conductive hearing loss, at 40 to 75 dB HL (dial setting) in 5 dB steps for participants with moderate conductive hearing loss, 50 to 85 dB HL (dial setting) in 5 dB steps for participants with moderately-severe sensorineural hearing loss, and at 70 to 95 dB HL (dial setting) in 5 dB steps for participants with severe sensorineural hearing loss. Participants were given two 5-minute rest breaks during this testing to improve co-operation and reduce fatigue, and the responses were scored by two scorers (the researcher and the isiZulu-speaking audiologists).

Data analysis

The PTA (the average hearing threshold for 0.5, 1 and 2 kHz) and zSRT scores for both participant groups were confirmed as meeting parametric assumptions by inspecting their histograms, box-and-whisker plots and Q-Q plots (data not shown). Agreement between the PTA and zSRT results for each participant group separately, for right and left ears separately, were assessed using intra-class correlation coefficient (ICC) analyses (Bartlett & Frost, 2008; Rankin & Stokes, 1998; Shrout & Fleiss, 1979) The strength of these ICC values was classified using the general (although arbitrary) guidelines reported by Landis and Koch (1977) of <0 indicating poor agreement, 0.01–0.20 indicating slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 almost perfect

agreement. Differences between the PTA and zSRT results for all participants as a single group, for right and left ears separately, were assessed using linear mixed model analyses. These analyses were conducted with threshold (dB HL) as the dependent variable, threshold method (PTA, zSRT) as an independent variable and fixed effect, participants as an independent variable and random effect (with unstructured co-variance type), and the intercept included in the model.

From session 2, logistic regression analyses were used to determine the psychometric function of each of the 28 words in the isiZulu SRT recording for each of the four subgroups in the hearing loss participant group separately. For each word in the isiZulu SRT recording, and for each participant subgroup, logistic regression equations were fitted, the slope of each model determined and the word threshold (50% correct) calculated. This analysis was not completed for the normal hearing participant group, as it had been completed on a separate sample of participants with normal hearing and reported elsewhere (Panday et al., 2009, 2018a).
All statistics were conducted using IBM SPSS Statistics, version 24, release 24.0.0.0 for personal computers.

Ethical considerations

Ethical clearance was granted by the Faculty of Health Sciences Human Research Ethics Committee of the University of Cape Town to conduct the study (clearance number: HREC 652/2012), and subsequent annual clearance was obtained during the data collection period. Informed consent was obtained from all participants prior to their participation in the study.

Results

For part 1 of the study, Table 8 shows the mean PTA and SRT values for participants with normal hearing and with hearing loss, while Table 9 shows the intraclass correlation coefficients (ICC) between the two threshold measures (PTA and zSRT) for each group by ear. The ICC values ranged between 0.67–0.88 for each ear for both single (single threshold method) and average measures (averaged between both threshold methods) for the normal hearing group, and were >0.95 for single (single threshold method) and average measures (averaged between both threshold methods) for the hearing loss group.

Table 8: PTA and zSRT measurements for right and left ears for participants with normal hearing (NH) and with hearing loss (HL)

	Pure Tone Average (dB HL)		Speech Reception Threshold (dB HL)	
	Right ear	Left ear	Right ear	Left ear
Normal Hearing (NH) group *(n = 100)*				
Mean	8.20	7.41	11.25	10.70
SD	5.48	5.57	5.28	5.45
Hearing loss (HL) group *(n =76)*				
Mean	56.53	58.53	58.42	57.37
SD	21.72	23.19	22.63	21.25

Table 9: Intraclass correlation coefficient (ICC, 2, 1) results for agreement between PTA and zSRT measures for participants with normal hearing (NH) and with hearing loss (HL)

	ICC (2,1) for normal hearing (NH) group *(n=100)*			
	ICC coefficient	95% CI	F test value	df
RE (NH)				
Single measure	0.78	0.69-0.84	99.0	75
Average measure	0.88	0.81-0.91	99.0	75
LE (NH)				
Single measure	0.67	0.55-0.77	99.0	75
Average measure	0.80	0.71-0.87	99.0	75
	ICC (2,1) for hearing loss (HL) group *(n=76)*			
	ICC coefficient	**95% CI**	**F test value**	**df**
RE (HL)				
Single measure	0.97	0.95-0.98	86.31	75
Average measure	0.99	0.98-0.99	86.31	75
LE (HL)				
Single measure	0.95	0.90-0.97	86.31	75
Average measure	0.98	0.96-0.99	86.31	75

*single measure refers to each threshold measure on its own and the average measure refers to average between the two threshold measures

Table 10 shows results of the linear mixed model analyses for both participant groups by ear for differences between the two threshold measures of PTA and zSRT, with a significant ($p<0.001$) difference between PTA and ZSRT being observed for each participant group.

Table 10: Linear mixed model analysis results for differences in thresholds by threshold measure (PTA versus zSRT) for participants with normal hearing and participants with hearing loss

Ear	Parameter	Estimate	Standard error	t	Sig	95% confidence interval
			Normal Hearing (NH) group *(n =100)*			
Right Ear	Intercept	11.10	1.57	7.057	<0.001	8.016 to 14.183
	PTA	-3.08	0.75	-4.10	<0.001	-4.575 to-1.60
	SRT	0				
Left Ear	Intercept	10.40	4.03	2.57	<0.001	1.284 to 19.515
	PTA	-2.91	0.79	-3.65	<0.001	-4.49 to-1.34
	SRT	0				
			Hearing Loss (HL) group (*n*= 76*)*			
	Intercept	60.87	2.50	24.34	<0.001	55.90 to 65.83
	PTA	-2.04	0.64	-3.18	0.002	-3.32 to -0.77
	SRT	0				
	Intercept	61.14	2.58	23.74	<0.001	50.02 to 66.26
	PTA	-3.03	0.64	-4.77	<0.001	-4.30 to -1.77
	SRT	0				

For part 2 of the study, Table 11 shows the psychometric function measures, and Figure 5 shows the psychometric functions for the participants with hearing loss by degree of loss. The psychometric data obtained from the present study's normally hearing participants are not reported here as they have already been reported in Panday et al. (2018a). The psychometric data obtained from another group of normally hearing participants using a precursor zSRT recording to the one used in the present study has also been reported by Panday et al. (2009).

Table 11: Psychometric function measures for the 28 isiZulu words for the participants by degree of hearing loss

	Mild conductive hearing loss			Moderate conductive hearing loss		
	Slope	Slope at 20% to 80% correct	Threshold	Slope	Slope at 20% to 80% correct	Threshold
Mean	4.92	3.15	37.09	5.26	3.37	54.46
SD	1.26	0.81	3.04	0.95	0.61	4.40
Minimum	2.88	1.84	30.40	3.65	2.34	45.59
Maximum	7.63	4.88	43.64	7.63	4.88	64.68
Range	4.75	3.04	13.24	3.98	2.54	19.09
	Moderately severe sensorineural hearing loss			**Severe sensorineural hearing loss**		
	Slope	**Slope at 20% to 80% correct**	**Threshold**	**Slope**	**Slope 20% to 80% correct**	**Threshold**
Mean	2.85	1.82	64.15	2.47	1.58	82.68
SD	0.41	0.26	5.21	0.53	0.34	5.93
Minimum	2.31	1.36	55.11	1.63	1.04	73.09
Maximum	3.75	2.40	79.00	3.68	2.35	94.68
Range	1.63	1.04	23.89	2.05	1.31	21.59

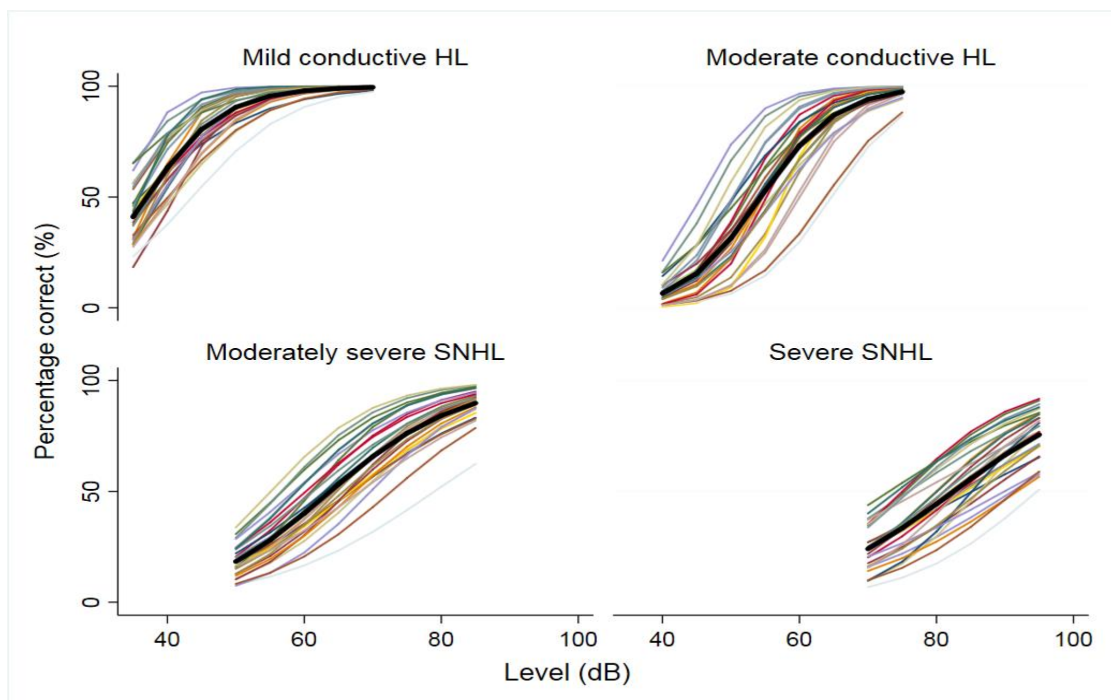

Figure 5: Psychometric functions for each of the 28 words in the isiZulu SRT test, for each participant group by degree of hearing loss. The bold, black line indicates the average psychometric function for all words for each psychometric function for all words for each participant group.

Discussion

The present study's results suggest the isiZulu SRT test is a valid measure of SRT in first language, adult speakers of isiZulu with and without hearing loss. The convergent validity of the zSRT test was supported by the very high agreement (ICC values ≥0.80) observed between participant PTA and zSRT values for participants with and without hearing loss, for both right and left ears. This support is drawn from this study's use of the zSRT and PTA scores as measures of the theoretical construct of hearing threshold. The degree to which the zSRT and PTA measures of this construct were related suggests high convergent validity. The zSRT results can reasonably be expected to show high levels of agreement with the established PTA results for first language, adult speakers of isiZulu with hearing loss when these tests are applied in clinical settings.

While the zSRT and PTA values were in very high agreement in this study's participants, mixed model analyses showed the zSRT values to be significantly ($p<0.005$) higher than the PTA values by an average of 2.9 to 3.0 dB for those with normal hearing, and by an average of 2.0 to 3.0 dB for participants with hearing loss. While statistically significant, these differences were not considered to be clinically significant, given that the minimum step size for presentation level used in this study was 5 dB, with differences of 6 dB or less between SRT and PTA having been taken to indicate agreement between those two measures for tests in several languages (Carhart, 1971; Han et al. 2011; Marinova-Todd, Siu & Jensad, 2011; Sreedhar, Venkatesh, Nagaraja & Srinavasan, 2011; Wang et al. 2007).

The concurrent validity of the zSRT test was supported by the psychometric functions for each word in the isiZulu SRT test being consistent with the degrees and types of hearing loss shown by the participants with hearing loss by subgroup, i.e., mild conductive, moderate conductive, moderately severe sensorineural, and severe sensorineural hearing loss. For increasing degrees of hearing loss, the psychometric functions showed the expected increases in threshold and decreases in slopes (Wilson & Carter, 2001) for the isiZulu SRT test words. Similarly, Wang et al. (2007) found that when participants with sensorineural hearing loss were evaluated with Mandarin bisyllabic words, the psychometric functions slopes were shallower compared to the results of their normal hearing counterparts.

Closer inspection of the psychometric functions of individual words in the isiZulu test across the four hearing loss groups indicates some variability in curves among individual words. For

example, the psychometric functions for the words 'wina' and 'minya' were consistently very shallow for all degrees and types of hearing loss. This suggests these words could be lowering the concurrent validity of the zSRT test overall, and their removal from the isiZulu SRT test wordlist could be considered.

Clinical implications

The present study's findings further support the potential use of the isiZulu SRT test as a valid and reliable measure of SRT in first language, adult speakers of isiZulu. This isiZulu SRT test has now been shown to contain words that are linguistically familiar and homogenous in audibility in isiZulu-speaking adults (Panday et al., 2007, 2009, 2018a,b) to produce reliable SRTs when used on adult speakers of isiZulu with normal hearing sensitivity (Panday et al, 2018b), and to produce valid SRTs when used on adult speakers of isiZulu with and without hearing loss (current study).

The present and previous studies reporting the development of this isiZulu SRT test also support the use of rigorous, systematic and multiple methods based on a unified validation framework to accumulate evidence for the validity and reliability of any new test. Attempts to use only one or two methods to confirm the validity and/or reliability of new tests should not be considered sufficient for such purposes. It is important to view multiple sources of evidence regarding the validity and reliability on new tests before they can be properly considered for clinical use.

Conclusion

The present study's results suggest the isiZulu SRT test is a valid measure of SRT in first language adult speakers of isiZulu with and without hearing loss, showing both convergent and concurrent validity when used in these populations. This study's findings are limited by its participants being predominantly recruited from the eThekwini Municipality and surrounding regions in KZN, although the test words did represent a central dialect of isiZulu. It could also be limited by its use of the SRT method described by Chaiklin and Ventry (1964, cited in Gelfand, 2001) with modification, which may not immediately generalise to SRT scores contained using other methods.

Declaration of interest

The authors report no declaration of interest.

Acknowledgements

The authors acknowledge Yusentha Balakrishna from the Medical Research Council of South Africa, Durban, South Africa, and Dr Asad Khan from the School of Health and Rehabilitation Sciences, The University of Queensland, Australia, for assistance with statistical analyses. The researchers also acknowledge Oticon foundation for research funding support.

Addendum

The following sections (6.1 and 6.2) relate to results for the normative study and the divergent validity study that were not part of the publication, the findings being linked to the validity of the isiZulu SRT test.

6.1 Results of the normative study

A total of 100 normal hearing participants (stratified according to age groups) were tested for pure tone audiometry and speech reception threshold testing using isiZulu and English (modified CID W 2) lists. In the paper above, the results for the normal hearing participants were described as a whole group, while in this section, the results obtained for the normal-hearing group will be presented according to each age group. These results provide the normative data for the isiZulu SRT test, with Table 12 providing the mean PTA and mean zSRT results with standard deviations for right and left ears.

The difference between the mean PTA and mean zSRT was 2,65dB and 3,39dB for the right and left ears respectively for the 18–30-year age group. For the 31–40-year age group, the difference between the mean PTA and mean zSRT was 3,07dB and 2,50dB for right and left ears respectively. Similarly, for the 41–50-year age group, the difference between the mean PTA and mean zSRT was 3,05dB and 2,83dB for the right and left ears respectively. A slightly higher difference between the mean PTA and mean zSRT was obtained for the 51–60-year age group, i.e., 3,83dB for the right ear and 4,71 dB. Overall, the difference between the mean PTA and the mean zSRT for all age groups was less than 5dB.

Table 12: PTA and zSRT measurements for right and left ears for participants with normal hearing (NH) according to age group

	Pure Tone Average		Speech Reception Threshold	
	Right ear	**Left ear**	**Right ear**	**Left ear**
18- 30-year group (N= 34)				
Mean	5.14	4.26	7.79	7.65
SD	3.58	3.71	3.73	2.89
31–40-year group (N= 26)				
Mean	7.69	6.92	10.76	9.42
SD	5.14	4.70	4.40	5.16
41–50-year group (N=23)				
Mean	9.34	8.69	12.39	11.52
SD	5.70	5.88	4.48	4.37
51–60-year group (N=17)				
Mean	13.52	12.64	17.35	17.64
SD	4.59	5.33	4.37	5.03

6.2 Results for the Divergent validity

The results of Objective ? of the study, a mixed model analysis, as indicated in Table 13, was considered to indicate if there was a difference between the mean PTA, mean eSRT and mean ZSRT scores.

Table 13. Mixed model analysis results for PTA average hearing threshold for 0.5, I and 2 kHz and isiZulu SRT (zSRT) and English SRT (e SRT) scores

	Estimate	Std. Error	df	t	Sig.	95% confidence interval
			Right ear			
Intercept	58.8	2.54	87.3	23.2	<0.001	53.8 to 63.9
eSRT	5.1	0.67	156.0	7.6	<0.001	3.8 to 6.5
zSRT	2.0	0.66	156.0	3.0	<0.005	0.7 to 3.3
PTA	0	0				
			Left ear			
Intercept	58.1	2.63	85.3	22.1	<0.001	52.9 to 63.3
eSRT	5.3	0.68	145.8	7.8	<0.001	3.9 to 6.6
zSRT	3.0	0.66	145.7	4.5	<0.001	1.7 to 4.3
PTA	0	0				

The linear mixed model analysis indicated that *for the RE analysis,* the mean SRT_E_RE (63.9 dB HL) > mean SRT_Z_RE (60.8 dB HL) > the mean PTA_RE (58.8 dB HL). While statistically significant, the differences in the means are not large. However, from a clinical perspective, the mean SRT _E_RE was 5 dB higher than the mean PTA_RE, while the mean SRT_Z_RE only ~3 dB higher than the mean PTA_RE. This was similar *for the LE analysis,* where the mean SRT_E_RE (63.4 dB HL) > mean SRT_Z_LE (61.1 dB HL) > mean PTA_LE (58.1 dB HL). While statistically significant, the mean eSRT was also 5dB higher than the PTA but for the mean ZSRT, it was 3dB higher than PTA.

Conclusion

This chapter presented the main findings for the convergent, divergent and concurrent validity of the thesis. It was found that the isiZulu SRT test is a valid measure of SRT in first language, adult speakers of isiZulu with and without hearing loss, showing both convergent and concurrent validity when used in these populations. In addition, this chapter presented results for the divergent validity findings, indicating that there was a significant difference between the isiZulu SRT and eSRT and PTA mean values. Furthermore, the clinical difference between the PTA and isiZulu SRT was marginally smaller than the e SRT and PTA. These findings suggest that the eSRT is different to the zSRT (although by a small margin) and to the PTA. The findings of this study has far reaching implications for both clinicians and researchers.

CHAPTER 7. DISCUSSION

7.1 Introduction

This chapter discusses the answers offered to the thesis aims and objectives. At one level, this thesis provides empirical and clinical evidence for the first reliable and valid isiZulu SRT test for KwaZulu-Natal Province, South Africa, as they have contextual implications for both clinicians and researchers. At another level, it offers guidance on how measurement-related elements [reliability and validity] within a systematic validation framework can be helpful to researchers who intend to validate similar speech audiometry and clinical tools. As this study has been presented in a hybrid format, the discussion for the relevant objectives have been included in each of the published articles presented in Chapter 4, 5, and 6. This chapter therefore provides a discussion of the main findings for each aim, expands on the clinical implications, and gives a more detailed focus on the application of the theoretical framework. It outlines the significance of the findings, indicates the limitations that may have impact on the study, and makes recommendations for practice, policy and research related to the topic.

7.2 Discussion

South Africa lacks reliable and valid speech audiometry tools in the indigenous languages of the country. In response to the clinical and contextual need for such tools, Panday et al. (2006, 2007, 2009) embarked on the task of developing an isiZulu Speech Reception threshold test for first language speakers of KwaZulu–Natal (KZN) Province, but lacked reliability and validity evidence, like many SRT tests in quiet both locally and internationally. There is a pressing need for such a test within the South African context, as Audiologists continue to use the English material, these being adapted non-standardised indigenous language tests, when assessing SRT in quiet for African language speakers (Panday 2006). This thesis therefore used a unified theoretical framework of validity (Chapter 3) to establish the reliability and validity of the isiZulu SRT test developed by Panday (2006). The findings are discussed in relation to this framework, with Figure 7.1 providing an illustration of the process followed in relation to the theoretical framework, and how each element related to each aim and objective of the study

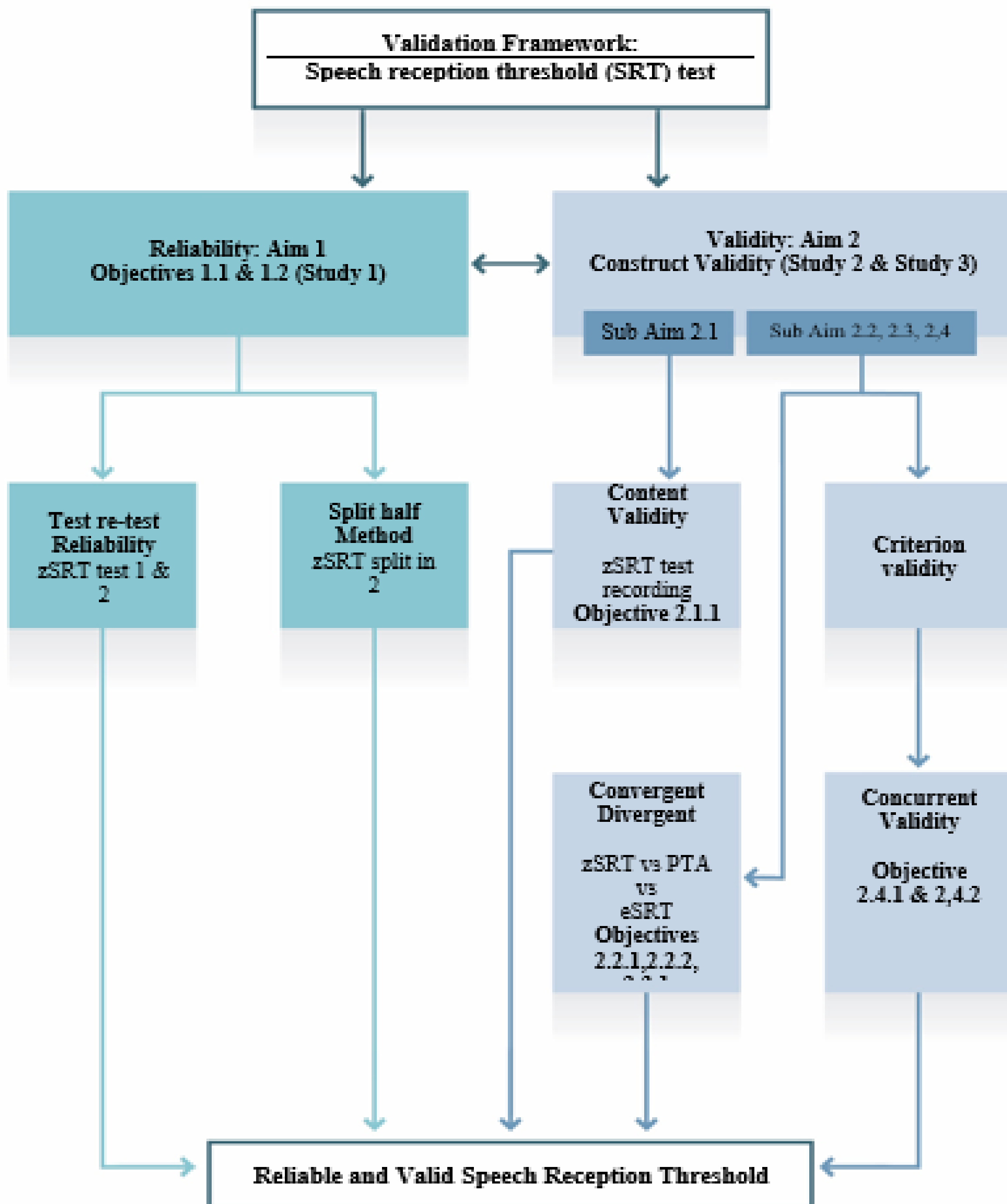

Figure 7.1 Validation framework for the isiZulu SRT test

7.2.1 Aim 1: To determine the reliability of the isiZulu test

Establishing the reliability of the isiZulu test entailed determining its test-retest reliability (Objective 1.1) and its internal consistency (Objective 1.2). This was done as reliability, also known as the internal and external consistency is essential and relates to validity (Mendel, 2008). In Study 1 of this thesis, evidence linked to the reliability of the test was foregrounded and included within a validation framework, as reliability and validity are interrelated. The

findings indicate that the isiZulu SRT test has high internal and external consistency, which renders it reliable and recommended for clinical practice. The high external consistency was established by the test-retest analysis results, which showed substantial agreement with the ICC (2,1) analysis, the range of difference scores on the Bland Altman analyses being within the ±5 dB variation in SRT scores expected within subjects in a clinical setting for the right ear stimulation used in this part of the study. These results indicate that the isiZulu SRT will remain consistent for isiZulu-speaking participants who experience no changes in their hearing over four weeks. These findings may have clinical significance for the local context, where monthly serial monitoring of SRTs and hearing may be required.

The favourable reliability findings in Objective 1 compare well with other studies that investigated test-retest reliability or external consistency of newly developed speech audiometry tests, for both in quiet and in noise (Kim et al., 2015; Lotfi et al., 2016; Neuman et al., 2012). Plomp and Mimphen (1979) emphasized that for speech audiometry tests to be used in clinical practice, there should be evidence that a high test-retest reliability exists in that target population. Test-retest reliability estimates, however, can be affected by various factors (Ostergaard, 1983), such as the learning effects and period between the test and re-test, number of test items, variability of the participant's age range and other linguistic factors that may influence the correlation.

While study 1 found a high test-retest reliability that compared well with the studies mentioned above, other speech recognition threshold tests in quiet (Caswell, 2013) and word recognition tests (Grange, 2013) returned greater than 0.85 and 0,90 for test-retest reliability for English spondaic word tests respectively. In these studies, participants were re-tested on the same day after a short rest period, with reports suggesting that when this occurs on the same day, or later (less than a week), the reliability of speech tests may be influenced by the well-known learning or memory effect (Maxwell & Satake, 2006). Learning effects are known as practice effects, which may result in improved performance as the test participants are more familiar with the task, test format, environment and materials. Therefore, in study 1, testing participants four weeks later mitigated any possible learning effect (Maxwell & Satake, 2006). The findings of the present investigation did not concur with other studies (Caswell, 2013; Grange, 2013), where the re-test was almost immediate (on the same day or within two weeks), returning almost perfect correlations.

Furthermore, participants were familiarized with the test words for study 1, which is part of most standard SRT procedures and a well-established step in the historical literature on SRT. Other studies removed familiarisation to avoid the memory and learning effect, while in this study, familiarisation with the test words was only done at the start of test 1, this approach having implications for methods used in reliability testing when speech audiometry tests are developed and validated.

Regarding the number of items in the test, the higher the number of test words or items, the higher the correlation coefficient, with narrower confidence intervals emerging after the test and re-test conditions (Kim et al., 2015). In the current study, 24 words were deemed sufficient for the SRT testing (Punch & Howard, 1985), with a substantial agreement being found. Although this is acceptable for a clinical test (Landis & Koch), the ICC for this study could have been better or different with a longer list of words. Nonetheless, selecting these 24 words was based on a careful process described by Panday et al. (2007, 2009) before the reliability study.

For the isiZulu SRT test-retest reliability, the age range of the participants was 17 - 51 years, the variability possibly influencing the overall results between tests 1 and 2, with correlations between them perhaps being lower when homogenous age groups are tested. Those studies with a broader variance among the participants (Caswell, 2013) are shown to return higher correlations, the age range in this study being wide, which may also have accounted for the substantial agreement for the different components of the study.

While the findings of Study 1 were positive and have implications for clinical use, the reliability data is based on participants with normal hearing, and did not include those with hearing loss. It is known that in clinical studies when test-retest reliability is investigated, the clinical stability of participants must be determined before the re-test is completed (Paiva et al., 2014). It is for this reason that the reliability investigations in this study were completed for the normative population. Including participants with underlying diseases, such as middle ear disease or progressive sensorineural hearing loss for the reliability of the study, would have impacted the clinical stability of the participants, thereby influencing the timing of the re-test. Future research, however, is recommended to investigate the reliability of the IsiZulu SRT test on participants with hearing loss, considering the importance of timing and learning effects. One such study noted that participants with hearing loss were considered and re-tested on the same

day (Caswell, 2013), the correlations being almost identical. In such a case, it is difficult to conclusively confirm if this was due to the use of heterogeneous groups and wide variance or the learning effect's impact.

Another limitation of Study 1 (reliability study) is that the SRT was established using the traditional method modified by Chaiklin and Ventry (1964) for establishing SRT. Other methods recommended for SRT testing in the literature may result in varying reliability findings for this test, with the ASHA (1988) method utilising a 2-step procedure compared to the 5-step method used in this study.

Finally, the interpretation of the reliability of this study was based on the Landis and Koch, and Bland and Altman methods, with the variation of +5/-5 dB being considered clinically acceptable after reviewing the literature (ANSI, 2004; ASHA, 1988; Caswell, 2013; Hallgren et al., 2006; Neuman et al., 2012). If a lower variation had been expected, these results would have to be interpreted differently, with most studies clinically accepted the variation of +5/-5 dB when SRT tests were considered for reliability testing.

At an empirical level, the evidence collected with regard to the reliability of the isiZulu test is useful for both clinicians and researchers, notwithstanding the limitations and criticisms presented above. Clinicians who are faced with the need to monitor patients regularly for their hearing ability, e.g. serial monitoring during ototoxicity management, may find this to be a helpful test, as it is reliable and consistent. In addition, the isiZulu SRT test may be used to assess the success of hearing aid use in adult patients, together with other tests within the audiological test battery. The overall clinical implications of the isiZulu SRT will be elaborated further in the chapter.

The findings of the reliability study (Study 1) have theoretical and research implications, with the reliability evidence also contributing to the overall validation of the isiZulu test. After careful examination of the Study 1 findings, researchers should consider collecting evidence by more than one method, as suggested by the theoretical framework. While researchers have and can independently investigate test-retest (external consistency) or split-half reliability (internal consistency), there are more substantial implications for the clinical use of these tests when the results are seen together. This is particularly important for clinical research in audiology when sample sizes for reliability studies can be small, and there is attrition and a loss of participants

during the test-retest study. Testing for both internal and external consistency of the test can improve its overall consistency and stability.

Additionally, there are some implications that relate to the choice of statistical methods to confirm the reliability of the test, this study utilising both the ICC (2.1) and the Bland Altman test. The result obtained from the ICC measurements indicated substantial agreement, while the Bland and Altman test of agreement also indicated a clinically acceptable variation. As indicated earlier, factors such as learning or memory effects, number of test items, and variability of participant characteristics in the sample, can influence the test–retest reliability estimate. Therefore, a single analysis method may not be sufficient to confirm the overall reliability of the clinical test.

The inclusion of the Bland Altman analysis methods, taking into consideration what is considered acceptable clinical variation for this construct, was required, this being different to studies that mainly used the Pearson r (Kim et al., 2015) to confirm the reliability of speech audiometry tests. Ng (2014) concurs that the choice of statistical measure to confirm reliability of speech tests is an important consideration, as they can improve the overall rigour of the study. Therefore, the evidence obtained for the reliability of the zSRT test through the combined analysis methods contributed to the empirical evidence regarding the overall validation of the test. It may be argued that this study contributes new reliability findings for the isiZulu test for clinical practice use, makes suggestions and provides further insight for improved methods that can be considered for other studies focusing on the reliability of newly developed tests.

While the newly developed zSRT test has been shown to be a reliable tool, more was needed to ensure that it is valid (Mendel, 2008). The evidence gathered for reliability contributes to the overall validation of the test and has implications for its validity. The next section provides the discussion of the empirical evidence related to the validity of the study.

7.2.2 Aim 2: To determine the construct validation of the newly developed isiZulu SRT test

Construct validity was viewed as an overarching concept in the validation of tests, as proposed by the adapted Messick Framework in Chapter 3. Several levels of measurement-related evidence were collected as part of the empirical evidence for the validation of the isiZulu test related to construct validity. Having discussed the findings for the reliability component of

Study 1, and the value of reliability studies within this framework, the findings specifically relating to the validity evidence are discussed both in terms of the empirical evidence and the theoretical and clinical implications.

Chapter 5, presents the results for Study 2, where content validity was examined by collected evidence from community members who were first language speakers of isiZulu about the actual recording of the test material. Community members were deemed the experts of the language as they use and engage with it daily. Ma et al. (2013) revisited a historical perspective on what constitutes the 'test material' in speech audiometry, and regarded it as more than just the actual words or word list. The recording and other acoustic properties of the words or test items are also important, these aspects having to undergo the same level of verification as the words themselves. Against this background, people who spoke isiZulu as a first language judged the recording for its naturalness in terms of intonation, tone, clarity, and how representative the recording was for the specific isiZulu dialect. This not only provides evidence for the content validity of the test, but also improves the overall face validity of the tests in question.

First language raters proved to be very useful in this study, as they were able to identify three words to be below the 'strongly agree' category for pitch or tone, and one below the 'strongly agree' category for clarity or articulation and naturalness/dialect (Panday et al., 2018). Study 2 also provided the added evidence on the content validity of the word recordings, which Panday et al. (2007) and other researchers had excluded. Other studies only utilized raters who were experts in the field to judge the test materials for cultural sensitivity (Nissen et al., 2005; Nissen et al., 2008), choice of words, as it relates to familiarity with the language (Nissen et al., 2005; Nissen et al., 2008), and naturalness of the words to the native speakers repertoire (Nilsson, 1994; Wong, 2008 in Theunissen et al., 2009) when content validity evidence was obtained.

Therefore, at the empirical level, Study 2's inclusion of first language speakers to rate the recordings was found to be valuable, as the linguistic and psychometric criteria employed by Panday et al. (2007; 2009) did not identify the problems that first language speakers were able to report on for three word recordings. The recordings revealed that the manner in which the speaker articulated these words could have influenced the pitch, tone, articulation and naturalness of these three-word recordings (Panday et al. 2018b). It was revealed on inspection of the actual recordings that the manner in which the speaker recorded the words may have

provided greater emphasis on their second syllable, which resulted in a slight change in the pitch and tone, particularly for the word ‘cinga’. The iambic meter where the emphasis is on the second syllable is different from the spondaic meter (equal emphasis on each syllable) that is present in the bisyllabic word recordings used in English language speech reception tests.

The merit of including first language speaker ratings while developing the isiZulu SRT test offered in this study is supported by suggestions that commonly used measures of linguistic criteria and homogeneity of audibility alone may be insufficient to ensure that the final test appropriately represents all aspects of the construct it is intended to measure (Maxwell & Satake, 2006; Nissen et al., 2005). Taking into account broader criteria, for example, the appropriateness of the word recordings, as evaluated by members of the communities in which the SRT test is developed, can provide direct and immediate feedback regarding the relevance of the word and test recordings being evaluated. The ratings serve to confirm the content and ecological validity of any final test recordings.

Pascoe, Rogers and Norman (2013) noted the need for researchers in South Africa to disclose the methods they use to develop and validate contextually relevant tests. Using community members or first language speakers as language experts when selecting word recordings for SRT tests is an additional method to include when developing contextually relevant materials, and is one of the recommendations emanating from this study, with researchers needing to broaden their opinions about who can be regarded as ‘experts’. For language-based studies, such as this one, while meeting the requirements for psychometric properties it is important to ensure the test’s rigor, considering the contextual factors to match its language may be as important.

The findings in Study 2 confirmed the importance of the methods used to establish the validity and reliability of new speech audiometry tests. The finding of this study indicate the importance of using multiple methods to accumulate evidence in a systematic manner, both for and against the new tests, and the importance of viewing them collectively to establish their true validity and reliability in the context in which they will be applied. To further support this claim, a discussion on the findings for convergent and divergent validity is provided. The results for these objectives were reported in Chapter 6, Objective 2.2.2 and 2.3.1 of Study 3 being to determine the convergent validity.

Convergent validity refers to how well the newly developed test compares or relates to an already established test or gold standard. In Study 3, the isiZulu SRT scores for individuals with both hearing loss and normal hearing were compared to the PTA (gold standard). A high agreement (ICC values ≥0.80) between PTA and zSRT values for those participants with and without hearing loss was found for both right and left ears. The high agreement between the zSRT and PTA suggests high convergent validity.

However, further analysis using mixed model analysis revealed that the isiZulu SRT scores were significantly higher ($p< 0.005$), by 2.9 to 3.0 dB for participants with normal hearing, and by an average of 2.0 to 3.0 dB for participants with hearing loss. These results may indicate that there is a statistical difference between the PTA and SRT, but that it is not clinically significant. It was also reported in other studies that differences of 6 dB or less between SRT and PTA having been taken to indicate agreement between those two measures for tests in several languages (Carhart, 1965; Han et al. 2011; Marinova-Todd, Siu & Jensad, 2011; Olariu et al. 2018; Sreedhar, Venkatesh, Nadaraja & Srinivasan, 2011; Wang et al. 2007)

Another possible reason for the statistical difference between the SRT scores and the PTA may be the language itself. Like other tonal languages, the isiZulu language has a higher concentration of low-frequency sounds, such as (bh, ph, ba, kha) (plosive sounds), which make up the phonetic composition of the low-tone verbs. It may be plausible to suggest that these low-frequency sounds can influence the recognition of these words, resulting in higher SRT scores than the gold standard. Van de Merwe and Le Roux (2014) also speculated about the perception of tonal variations in Nguni languages, such as isiZulu, especially for participants with hearing loss. Similarly, Han et al. (2011) reported that the Koren reference SRT scores were influenced by the high concentration of low-frequency energy compared to English spondee SRT and the gold standard. Like most tonal low energy stimuli, bisyllabic words are not equally stressed in comparison to spondees, which are equally stressed, and at low threshold levels, the tonal variation is more difficult to identify than spondees (Han et al., 2011; Maronova-Todd et al., 2010). Marinova-Todd et al. (2011) further confirmed that the low-frequency components in tonal languages, such as Cantonese, might indicate marginal differences with the PTA, as its acoustic spectrum was also based on English sounds. While empirical evidence exists for the convergent validity of the newly developed test, based on the high ICC correlations and more detailed linear mixed model results, considering the linguistic factors are also essential for the overall validation, particularly for tonal languages.

Therefore, the divergent or discriminant validity evidence also needs to be obtained for this language-based study. When discriminant validity evidence is gathered, there a difference in scores between the new test and some other measure that does not reflect the theoretical construct. In this study, the zSRT and the eSRT were compared to the PTA, the intention being to determine whether there was a difference between these comparisons. Clinically, there was a 0-3dB difference noted between the zSRT and PTA measures, and a 5dB difference between the eSRT and PTA, which is within the acceptable clinical difference range, while statistically, there was a difference between the eSRT and zSRT and the PTA ($p < 0.001$). Further analysis of the results indicated that the difference between the eSRT and PTA compared to the zSRT and PTA varied. Although there was high consistency between all these measures, a statistical difference was noted, despite the clinical difference being negligible. These findings have important clinical implications that will be elaborated on later in this chapter. Despite the overall better performance and closer relationship between the zSRT and the PTA in comparison to eSRT and PTA, the high agreement between all measures, including the eSRT, was unexpected, and an interesting finding with plausible explanations.

Most of the participants selected for this study were from the eThekwini Municipality due to logistical reasons and the availability of volunteers, despite the wide range of recruitment strategies mentioned in each of the papers published. Although isiZulu is the first language for most people in eThekwini (68%) (Census, 2022), it is an urban metropolitan area, and participants may have been exposed to English in the media. Younger participants in the 18- 25 and 26-40 age groups may also be more familiar with the English language through social media and education, which may have impacted their performance more than expected in the English test in this study. Participants from other peri-urban and rural contexts may have performed differently on this test. While careful selection of participant groups for such a study is widely mentioned in the literature (Hapsburg, 2002), finding people who meet the specific language characteristic is not always a simple matter. These findings are therefore a good starting point for clinical practice, but it may also suggest a need for ongoing refinement and further field-based validity testing of newly developed tests in various clinical contexts.

This is the first recording of the spondee words in English with the same dialect as those of the participants, familiarity with this dialect of English possibly having improved the test performance. The VU peaking of this English word recording was also great at +/- 2 dB, although during the English testing, the necessary adjustments were made on the audiometer,

as described and recommended in Panday (2006; 2007), as was the original recording, described in Panday (2006). These words did not undergo any adjustments compared to the rigorous adjustments that the isiZulu recording underwent. There was still some variability in the VU meter peaking (some words peaked at greater than +/- 2dB) for the English word recording, which would have influenced the test performance. DiBeradino et al. (2010) confirmed that the higher gain between the calibration tone and the CD recording may influence the speech outcomes. This could be another reason for the unexpectedly close relationship between the English SRT and PTA.

The findings however confirm that there is indeed a difference (although marginal in this study) in performance when English and Zulu words are used for divergent validity, these findings may have been due to many methodological factors mentioned above. Therefore, the findings also suggest that in confirming the validity of new tests, one type of validity finding should not be viewed in isolation, as the design elements may have an impact on the outcome. These findings confirm that having multiple levels of validity evidence contributing to the overall construct validity is required for complex speech audiometry tests. The next level of empirical evidence was for the concurrent validity of the newly developed test.

This study demonstrated the concurrent validity of the zSRT test, as the psychometric functions for each word in the isiZulu SRT test aligned with the levels and types of hearing loss present in the study's participants, categorized into subgroups: mild conductive, moderate conductive, moderately-severe sensorineural and severe sensorineural hearing loss. As the severity of hearing loss increased, the psychometric functions showed expected patterns, with thresholds increasing and slopes decreasing, in accordance with Wilson and Carter's (2001). The results are consistent and reflect the typical characteristics of individuals with hearing loss when listening to speech. Wang et al. (2007) also reported that when participants with sensorineural hearing loss were evaluated with Mandarin disyllabic words, the psychometric functions slopes were shallower than the results of their normal hearing counterparts. Although a limited body of evidence exist for such studies in quiet on individuals with hearing loss, the findings in this study do concur with previous studies (Boothroyd, 2008; Wang et al., 2007). Furthermore, as reported in Study 2 and in Panday et al. (2009), the psychometric functions of individuals with normal hearing were found to be at a slope of 5.98%/ dB., which were far steeper than these obtained for the various individuals in the hearing loss groups. This suggests that the isiZulu SRT word list is considered suitable for use for individuals with hearing loss.

The psychometric functions of individual words in the isiZulu test across the four hearing loss groups were further examined, with variability in some of the curves. Those for the words /wina/ and /minya/ were consistently shallow for all degrees and types of hearing loss, which suggests that they could be lowering the concurrent validity of the zSRT test. The removal of these words from the isiZulu SRT test wordlist can be suggested in order to improve its overall validity. Psychometric functions have historically been used to measure the overall validity of the newly developed tests both in quiet and noise. While this was not a complete item analysis of each word, it provides an extension to the use of psychometric functions to identify problematic words for individuals with hearing loss, which must be removed to improve the overall validity of the test. The implication of this approach is that the final isiZulu list, after collective examination of the evidence, would be is suitable for clinical practice for individuals with and without hearing loss.

The implications of these findings also suggest that item analysis of the newly developed words should be included as an additional layer to the overall validation framework. These recommendations concur with Ng et al. (2016), who utilized item response theory and RASCH modelling to validate new Cantonese Speech Audiometry tests for children. The findings of the study suggest that the isiZulu test is a reliable and valid test, however, the application of the study findings for both clinicians and researchers is worth exploring and discussing, the ensuing discussion presenting the overall clinical implications.

7.3 Clinical Implications

The zSRT test is expected to show high levels of agreement with the established PTA results for first language, adult speakers of isiZulu with and without hearing loss when these tests are potentially applied in clinical settings. Local audiologists searching for a contextually and linguistically relevant isiZulu SRT test to establish SRT with first language adult clients could utilizes this zSRT test in recorded format. The isiZulu SRT test has undergone several levels of development in the earlier work by Panday (2006) and Panday et al. (2007; 2009), with the current study having provided the evidence for the reliability (Panday et al. 2018b), content validity (Panday et al 2018a), as well as the convergent validity and concurrent validity (Panday et al., 2020). In addition, this study also provided normative data for the isiZulu SRT test across various age groups, as well as divergent validity evidence to show that when compared to the agreement seen with the English words, the isiZulu words returned better results. The expected zSRT and PTA (using the 3-frequency average) is likely to be less than 5dB, with the empirical

evidence provided in this study having value for clinical settings in KZN. Based on the findings of the study, further field-based testing is also suggested in clinical settings, which may attract individuals from the peri-urban and rural settings, as the study may be limited by choosing participants mainly from the eThekwini Municipality.

The dialect used in this word list recording is a central dialect of isiZulu and may have wide application. This test, despite the limitations mentioned, has significant value, and adds to the scarce body of clinical resources of contextually and linguistically relevant resources in South Africa. The isiZulu SRT test may be used as part of the behavioural test battery to confirm the pure tone testing for various differential diagnosis e.g. when functional hearing loss and pseudo hyperacusis must be confirmed in the clinical context, and when hearing of adults is diagnosed for both conductive and sensorineural hearing loss. The clinical settings in KZN, and South Africa, has for a long time relied on materials that were informal, non-standardized and translated, which lacked any research evidence to make important clinical decisions. This test offers clinicians the first African language test designed on carefully structured parameters and criteria for SRT that is aligned with the structure of the language, notwithstanding the limitations of the study.

However, from the unexpected findings of this study for eSRT, it may also seem tempting for clinicians to consider the eSRT test used in this thesis. The researcher is more cautious with this recommendation, mainly because the majority of the population who speak isiZulu as a first language may not share the same level of functional proficiency or familiarity with the English words to the participants in this study, for reasons already mentioned above. Only 41% of African adults indicated in a National Household (NIDS) survey that they are functionally proficient in English (Posel & Zeller, 2010). However, improved levels of proficiency in the English language by second language speakers may show improved SRT scores when the target language is English (Killman et al., 2014). The conclusions made by Killman et al. (2014), however, still firmly support that while SRT performance is good in such a target group, it is still better to choose a test that corresponds to *the native or mother tongue* of the listener, mainly due to linguistic or language factors, as discussed. *Language dominance* also continues to be the most important factor to consider when selecting a speech test, such as SRT. Shi and Sanchez (2010) also argue that although language proficiency, increased immersion (as seen in the present study) and age of acquisition are important considerations for selecting speech perception tests, *language dominance* is probably the most important factor to consider. Thus,

utilizing the newly developed isiZulu SRT test for first language speakers of isiZulu is a more suitable option.

For audiologists to consider the use of the modified English test, they would first have to establish the English proficiency of their clients, which would usually be done using self-reporting methods i.e. the client is asked to self-rate their own levels of proficiency. The question that can be raised for this context relates to the appropriateness of a method for selecting a speech reception threshold test. Relying on self-reporting on language proficiency in a clinical setting for the whole of KZN may result in overestimating their competency in relation to the task expected. Posel and Zeller (2010) confirmed that 41% of African adults in South Africa generally report functional proficiency in English, they also warn that these results may be influenced by over-reporting.

In addition, within the South African context, the indigenous languages, such as isiZulu, have not shared the same language status as the hegemonic languages, such as English. Some speakers of indigenous languages would be either hesitant or may overestimate their proficiency in English when asked to do so in a clinical setting. This may be mainly due to the historical power placed on languages such as English, where proficiency implied a higher level of education and social status. Many years post democracy, these deeply rooted effects of apartheid and colonialism may continue to influence daily communication and interaction. The considerable disparity between the number of first language English and Afrikaans speaking audiologists compared to the large population of people who speak an African indigenous language, such as isiZulu as a first language, exacerbates this imbalance. With the continued 'language-power imbalance' between therapists and patients or clients, self-reporting alone as a method to determine the levels of language proficiency when selecting a test within the KZN context may not be a suitable or culturally sensitive option, particularly for clients in rural areas. The newly developed isiZulu SRT test would therefore be a more reliable and valid measure than the modified version of the English test to select for SRT testing with first language isiZulu speakers.

The findings of this study therefore also have implications for the community of isiZulu speakers in KZN and the rest of the country. For a long time, despite strides made in the country to improve the status of health care services for underserved communities, the availability of appropriate resources and test materials were scarce. This thesis provides assurance that a test

exists that is familiar to the isiZulu language, and that the services offered through this SRT assessment is culturally sensitive and fair. The study findings also encourage clinicians to rethink how tests are selected to serve the multilingual and multicultural communities. Clinicians are encouraged to consider tests that are based on evidence-led research, as the science of speech audiometry research has shifted and continues to grow. It has shifted beyond the collection and development of words that were mainly tested on normative sample only. This study has shown that despite speech audiometry test development and validation being complex, there is a need for multiple levels of carefully designed steps before the test can be used in practices. This study suggests that clinicians also have a role to play in the ongoing validation of such materials through clinical field testing in the local context, which should be considered within the overall validation framework.

7.4 Theoretical implications

Based on the work of Messick (1995), this thesis considered a framework that was unified, and combined reliability and validity evidence, the intention being to develop a framework with a structured approach for developing and validating Audiology tools. These components are essential, as the outcomes of the test have implications for the people being tested, and could affect their overall hearing diagnosis and the accuracy of test procedures used. This means that every effort needs to be made to ensure that the test results accurately reflect the construct of what is being assessed.

Construct validity is in the center of the theoretical framework used in this thesis, as it was viewed as an overarching type of validity. The preceding discussion has shown how, through the study findings from each of the measurements related elements (both reliability and validity), the evidence contributes to the overall construct validity of the isiZulu SRT test. The findings presented for the various study objectives confirms that the evidence for such a study as a whole cannot be viewed in isolation, but that each piece must be viewed collectively. In the original description of validity theory, Messick highlighted the importance of the inferences that can be made from test scores rather than the scores itself. This study indicated that for some of the measurement related elements, such as reliability estimates, there was substantial agreement, but that the results from the Bland Altman method suggests that there may be variability, which was acceptable clinically. This was a good example that examining the reliability scores on their own may not be sufficient to confirm reliability, this approach having been used in other studies.

However, the study's findings suggest that additional factors should be considered as contributing to an extended framework for validating speech audiometry tools throughout this study. This extended framework (Figure 7.2) therefore brings together into one methodology aspects that relate to validity and reliability but are also influenced by contextual factors. The study findings in this thesis show the impact that these contextual factors have on the overall validation process, as illustrated in Figure 7.2, making it worthwhile to include them within the framework. Whereas in the initial framework suggested for this thesis, reliability and validity appeared as interrelated measurement variables, the extended framework now suggests that the contextual factors intersect both these variables, highlighting their importance in validating such tools. These contextual factors include the influence of the language of the test, choice and type of participants, recruitment of these participants, and the inclusion of community members as experts as part of the validation process. These contextual factors are further elaborated:

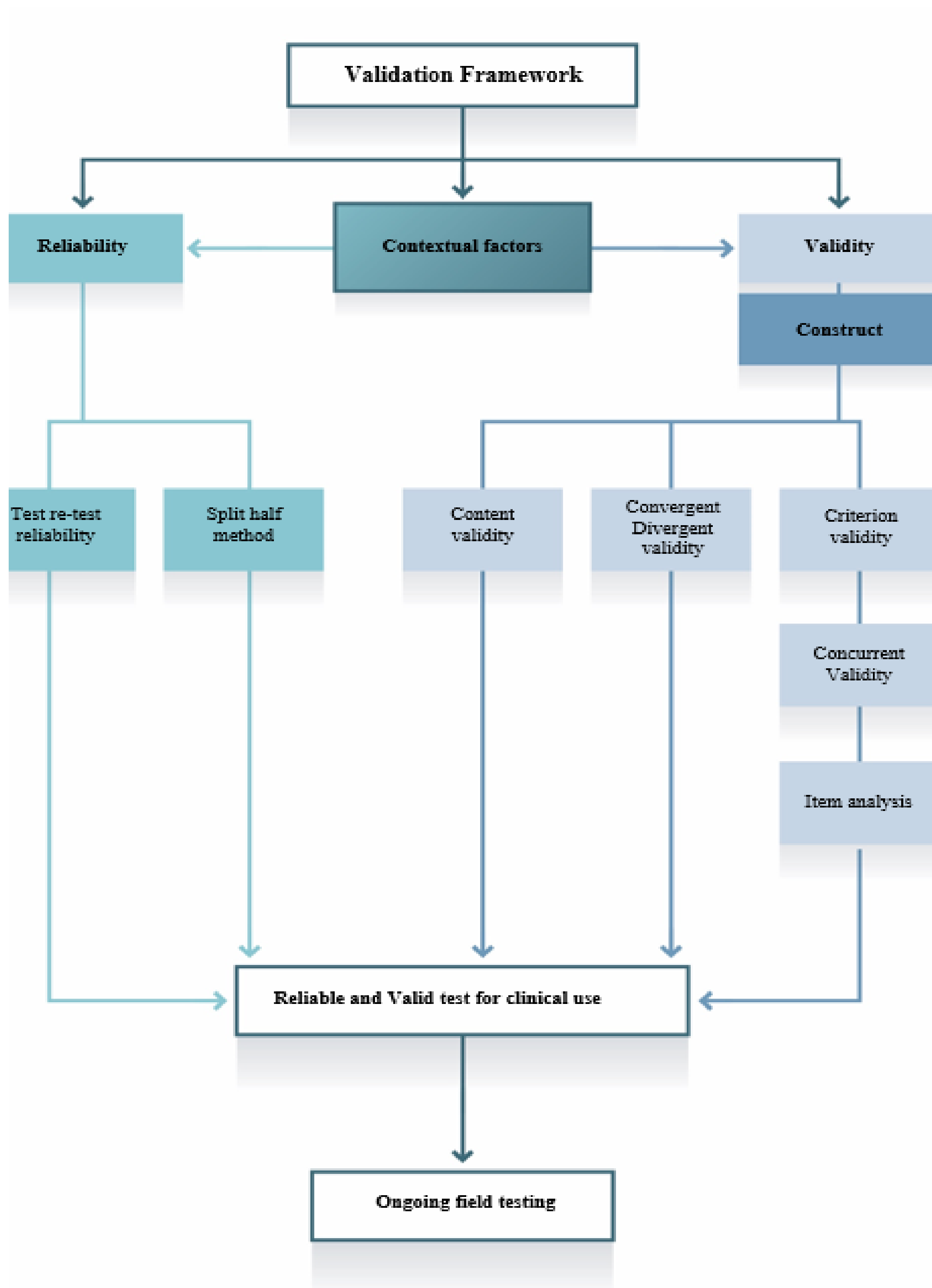

Figure 7.2 Extended Validation framework for speech audiometry tests in indigenous languages.

a. Implications for indigenous language and language influences in the validation process

In gathering evidence systematically and following the structured theoretical framework, what has emerged from the findings is that the complexity and richness of indigenous

languages does have an impact on the empirical interpretation of the scores. The convergent validity in this study was high when the ICC (2.1) results were examined, but when the mixed model results were added, there were significant differences between PTA and zSRT, these not being clinically significant. The phonetic composition of the words differs between languages, which may influence comparison to typical three frequency averages using pure tones. Historically, the similarity in scores with the pure-tone audiogram (PTA) has been used as an indicator of a valid speech reception threshold (SRT) and the absence of significant difference between the two variables. However, for tonal languages such as isiZulu, Korean and Mandarin, as discussed above, the scientific lens must be adjusted to accommodate a broader range of agreement, with a greater acceptance of marginal differences

While the framework provides a rigorous and systematic approach to validation, to fully confirm the construct of speech reception, it is essential to adopt a broader, more ongoing perspective on validation. The language in which such tests are developed plays an important role and warrants closer examination. While the findings of this study are undoubtedly positive and contribute novel evidence to the literature, confirming the first reliable and valid isiZulu Speech Reception Threshold (SRT), it is important to acknowledge the limitations of an only positivist paradigm in the context of language-based tests. A more interpretative paradigm may be necessary to address the complexities of language and its impact on test applicability, this being further expanded when the value of community members as experts within the context of language based tests is taken into consideration.

b. Experts

In this case of the isiZulu SRT, the evidence gathered for content validity challenges the traditional definition of 'experts', shifting this to community members rather than linguists only. This shift is due to the community members' ability to identify problematic words in the recordings, which may have otherwise been included based solely on standard psychometric functions and linguists. For example, the word *cinga* was rated as "strongly disagree," indicating that it did not align with the expected pitch and tone of the language. However, when the psychometric functions for this word were analysed across all hearing loss groups, the slope of the curve was found to align with the type and degree of hearing loss. Specifically, the curve was steep for the mild conductive hearing loss group (6%/dB), decreased slightly for the moderate conductive hearing loss group, and was much shallower for the moderately severe (2.75%/dB) and severe sensorineural loss

groups (2.95%/dB). Viewed in isolation, these psychometric results would suggest that the word is appropriate for inclusion. When the finding was viewed collectively, taking into consideration the psychometric properties and the perspectives of community members, there is evidence to indicate that the concurrent validity may be lowered, and the word may not have been recorded with accurate pitch, tone or clarity. These words would therefore have to be omitted from the final list (Appendix L)

c. Participants

The various factors discussed in this study, which pertain to participant recruitment and the complexities of selecting and retaining participants, highlight significant challenges in research and clinical settings with diverse populations, such as those in KZN. These challenges are not limited to linguistic diversity, but extend to cultural, socio-economic and health-related factors, including the high burden of disease in the region. Such complexities necessitate a reconsideration of how these factors influence the reliability and validity of new diagnostic tests and clinical research. As observed in this study, for multilingual environments such as KZN, recruiting monolingual participants for language-based studies can be particularly challenging. As exposure to multiple languages may affect the overall validity of the tests, the language used in the test, along with its broader socio-cultural context, plays an important role in participant recruitment, suggesting the need for a more inclusive framework for test validation. This broader approach should acknowledge that test validity outcomes cannot be considered a single reality, but rather a process influenced by multiple factors and perspectives.

The traditional view of test validation, which relies on a normative sample for clinical practice, is also questioned in this study, and highlights the need to incorporate additional contextual factors into the validation process, which should be ongoing rather than static. Messick's work on the consequential aspects of test validity underscores the importance of considering these factors. Therefore, the framework for test validation should be strengthened by explicitly incorporating contextual factors, ensuring a more comprehensive approach to test evaluation. The importance of context and cultural factors concurs with the findings and recommendations made by Abrahams and Khoza-Shangase (2025) and Khoza- Shangase and Mophosho (2018).

A counter view may argue that, practically, validation studies of this nature are unrealistic and idealistic, and are what should be considered at the very minimum for potential

clinical use. At a minimum, the extended framework suggested is regarded as having sufficient elements to confirm the reliability and validity of the new test. Considering each element within the framework in an integrated way confirms that a test has the potential to be used in clinical practice. However, for the framework to be all encompassing, the additional contextual and design related elements must be integrated within it. Therefore, based on the findings of this study, the framework used is further extended, with Figure 7.2 providing the extended version for validation of new speech audiometry tests in indigenous languages.

7.5 Significance of the study

The thesis focused on the validation of the first zSRT test in KZN, South Africa, and has significance both in terms of clinical and research advancements and social impact. This thesis contributes to improving the diagnostic accuracy when first language isiZulu speakers' hearing is assessed. Using a linguistically-matched speech audiometry test will ensure accuracy for people who speak that language, as the use of English language tests may miss the linguistic nuances due to the difference in phonetic and linguistic structure of isiZulu. This thesis also contributes to cultural relevance in audiology testing in South Africa, as traditionally, speech audiometry tests utilize words or stimuli that are unfamiliar to indigenous language speakers.

The thesis promotes indigenous languages, such as isiZulu, in the country, and enables it to be recognized and intellectualised. isiZulu speaking people have often faced significant health disparities, including limited access to healthcare and culturally appropriate tools. This study addresses one such gap by providing a diagnostic tool that is tailored to the needs of the speakers.

The validation of the zSRT test therefore contributes to health equity by ensuring that isiZulu language speaking people, regardless of their language or cultural background, have access to precise hearing evaluations and subsequent treatments. This study represents a step toward addressing historical and ongoing inequities faced by indigenous people in accessing culturally competent healthcare. It recognises the constitutional rights of people to have their health services in their first language, which respects their languages and cultural practices.

Of further significance is that this study involved the participation of members of the community in its development and validation process. This inclusion is essential for ensuring

that the test truly meets the needs of the community and respects their preferences and cultural practices. By employing a unified validation framework, this study may serve as a model for developing similar speech audiometry tests in other indigenous languages. It could be considered for creating linguistically and culturally appropriate tools for other marginalized or linguistic groups around the world.

The methodology and framework used in this study could potentially be adapted and applied to other areas of healthcare, such as cognitive or psychological assessments, where language and cultural considerations are also important. This study may start new conversations within the field of audiology about how best to approach the validation of speech audiometry tools. It advocates for a more inclusive and broader approach to validation of such tests. Finally, if the study’s findings and methodologies are accepted, it may lead to new educational paradigms in audiology, encouraging future practitioners and researchers to consider linguistic contextual factors within the validation framework, The thesis therefore contributes the first reliable and valid African indigenous language SRT test that was tested utilising a systematic validation framework, this being worth considering for the future.

7.6 Limitations

The study may have been limited in the following ways:

- The participants were mainly selected from central areas of KZN, despite many attempts to recruit from the wider community.
- The sample size for the reliability study was small and limited to normal hearing individuals only.
- The use of the Chaiklin and Ventry (1964) method for establishing the SRT could suggest different outcomes if other methods are utilized.
- The modified English language test was not adjusted technically as much as the isiZulu SRT test.

7.7 Recommendations

The study has provided a foundation for research, policy and practice, with the following recommendations being made:

- Research into the development and validation of speech audiometry tests in indigenous languages following a systematic and ongoing approach to validation.

- Research to test the reliability of the zSRT test on individuals with hearing loss, taking the learning and memory effects into account.
- Further field testing of the zSRT with the various dialects of isiZulu in the country.
- Improve the current zSRT test by including a more detailed item analysis.
- Develop and validate a word recognition test in quiet and noise using this framework for isiZulu Speakers.
- Develop and validate a paediatric zSRT and word recognition test.
- Strengthen the National Policy with respect to minimum standards regarding the development and validation of tools, the absence of which means that their consistency and applicability to the clinical population may be questioned.
- Refine existing SRT and speech audiometry test that have minimal evidence for reliability and validity to include more measurement-related elements and improve their design-related elements.
- Test the theoretical framework for its applicability to develop and validate speech audiometry tests in other languages.

7.8 Summary

The isiZulu Speech Reception Threshold (zSRT) test has been demonstrated to be both reliable and valid for use in populations with hearing and hearing impairments, this study providing substantial evidence across multiple levels, confirming its reliability and validity. It also highlights the importance of a systematic approach to validation, with each level of evidence contributing to the overall rigor of the study and enhancing the construct validity. The findings establish the zSRT test as a reliable and valid tool for clinicians, although some limitations should be acknowledged. Furthermore, the methods and framework outlined in this study may be useful for researchers developing and validating similar tools in other languages.

The study raises important considerations regarding the relative importance of different types of validity evidence in confirming a test's validity, and prompts reflection on the degree of validity required before a tool or test can be deemed suitable for clinical use. In light of these findings, validation studies should be carefully planned, with particular attention to the selection of a validation framework. Reliability evidence must be an integral part of any validation study, as must the careful selection of measurement-related elements contributing to construct validity, such as content, convergent and concurrent validity. This study also recommends the need for further item analysis to fully assess the test's validity and reliability.

Additionally, contextual factors, such as the selection of test participants, the involvement of community members in the validation process, and the impact of language on test selection and psychometric properties, must be carefully considered. To draw definitive conclusions about discriminant validity, the participant profile needs thorough examination, including their functional proficiency in both test languages. Moreover, geographical location and the influence of dominant languages must be considered, particularly in multilingual contexts. The generalizability of the results across regions with multiple dialects and the influence of the dominant language should be further explored.

7.9 Conclusion

The aim of this study was to assess the reliability and validity of a newly developed zSRT test for both hearing and hearing-impaired individuals in KZN. This thesis reports on the findings from three primary studies that examined the test's reliability and validity, utilizing a systematic integrated validation framework. The results indicated that the zSRT test is a reliable tool for the normative population and demonstrated collective evidence of overall validity for both hearing individuals and those with hearing impairments in KZN. In this study, reliability and validity were evaluated in an integrated manner, with each measurement-related element contributing to the overall construct validity of the test. The thesis thus presents a viable tool for potential clinical use in KZN, particularly for first language isiZulu speakers, although it acknowledges certain limitations.

The view of validation in this study as an ongoing process suggests that additional studies within clinical settings may be necessary to further refine this tool. The thesis also provides a theoretical framework and appropriate methodologies that will be valuable to both clinicians and researchers pursuing future test development. Based on the findings of this study, as well as prior work related to the development of the zSRT test, a key implication for researchers is the importance of adopting alternative perspectives on validation. Researchers are encouraged to employ a holistic, comprehensive and multi-level approach to evidence gathering before a test can be considered fully validated. At a minimum, validation studies should include reliability testing, content validity (with input from community experts), and assessments of convergent and concurrent validity, which should be viewed collectively to confirm the overall construct validity of the test.

This study introduced an extended framework for future validation studies in speech audiometry and test development, an important feature being the emphasis on contextual factors, positioning them as key elements in the validation process. In addition to its methodological contributions, this thesis offers the first systematically validated SRT test in an African tonal language, both locally and globally. The development of tests in local languages is essential for ensuring that services are appropriately tailored to the needs of the populations they are intended to serve. The use of standardized, culturally fair assessments enhances the quality of service delivery, thereby improving hearing healthcare and overall health outcomes. This study lays a strong foundation for future research in speech audiometry test development, both within South Africa and across other regions in Africa, where multiple indigenous languages are spoken.

REFERENCES

Abrahams, F., & Khoza-Shangase, K. (2025). South African Speech-Language Therapy and Audiology students' experiences of academic and clinical curriculum transformation. *South African Journal of Communication Disorders*, 72(1), a1086. https://doi. org/10.4102/sajcd. v72i1.1086

Aleksandrovsky, I. V., McCullough, J. K., Wilson, R. H. (1998). Development of suprathreshold word recognition test for Russian-speaking patients. *Journal of American Academy of Audiology*, *9*(6), 417-425.

Altman, D.G. (1991). Some common problems in medical research. In D.G. Altman (Ed.), *Practical statistics for medical research* (1st ed., pp. 398-403). London: Chapman & Hall.

Andrade, A. N., Sanfins, M. D., Skarzynska, M. B., Skarzynski, P. H., & Gil, D. (2023). Temporal ordering and auditory resolution in individuals with sensorineural hearing loss. *International Archives of Otorhinolaryngology, 28*(1), e122-e128. https://doi.org/10.1055/s-0042-1759748

American National Standards Institute. (1977). *Standard criteria for permissibility of ambient noise during audiometric testing (ANSI S3- 1977)*. Author.

American Speech-Language-Hearing Association (ASHA). (1988). Guidelines for determining the threshold for speech. *ASHA, 3*, 85–88.

American National Standards Institute. (2004). *Specification for audiometer (ANSI S3.6-2004).* American National Standards Institute.

American Speech-Language-Hearing Association (ASHA). (1990). Guidelines for screening for hearing impairments and middle ear disorders. *ASHA*, *30*, 85–88.

Ballachanda, B. (2001). Audiological assessment of linguistically and culturally diverse populations. *Audiology Today*, *3*, 34–35.

Bartlett, J. W., & Frost, C. (2008). Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstetrics and Gynecology*, *31*, 466-475.

Bellis, T. J., & Bellis, J. D. (2015). Central auditory processing disorders in children and adults. *Handbook of clinical neurology*, *129*, 537–556. https://doi.org/10.1016/B978-0-444-62630-1.00030-5

Bland, M. (1987). Clinical measurement. In M. Bland (Ed.), *An introduction to medical statistics* (pp. 265-273). Oxford: Oxford University Press.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet, 1*(8476), 307-310.

Borg, E., Wilson, M., & Samuelson, E. (1998). Towards an ecological audiology: Stereophonic listening chamber and acoustic environmental tests. *Scandinavian Audiology*, *27*, 195-206.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 135–170). IAP Information Age Publishing.

Cope, A.T. (1982). *A comprehensive course in the Zulu language.* Durban Howard College, University of Natal.

Caswell, K.L. (2013). *Test-Retest reliability of Speech Recognition Threshold material in individuals with a wide range of hearing abilities*. Unpublished thesis, Brigham Young University, Utah, USA.

Carhart, R., & Jerger, J. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders*, *24*, 330–345

Carhart, R. (1971). Observations on relations between thresholds for pure tones and for speech. *Journal of Speech and Hearing Disorders*, *36*(4), 476–83.

Chaiklin, J. B., & Ventry, I. M. (1964).Spondee threshold measurement: A comparison of 2dB and 5dB methods. *Journal of Speech and Hearing Disorders*, 29, 47–59.

Craig, C.H. (1997). Spoken language processing. In G.D. Chermack, & F.E. Musiek (Eds.), *Central auditory processing disorders: New perspectives* (pp. 71-90). Singular Publishing Group.

Cronbrach, L. J. & Meehl, P. E. (1995). Construct Validity in Psychological tests. *Pychological Bulletin, 52* (4), 281-302.

Dellinger, A. B., & Leech, N. L. (2007). Towards a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, *1*, 309-332 Retrieved from http://online.sagepub at University of Cape Town March 30 2011.

Di Berardino, F., Tognola, G., Paglialonga, A., Alpini, D., Grandori, F., & Cesarani, A. (2010). Influence of compact disk recording protocols on reliability and comparability of speech audiometry outcomes: acoustic analysis. *The Journal of laryngology and otology*, *124*(8), 859–863. https://doi.org/10.1017/S0022215110000782

Fox, A. (2000). *Prosodic features and prosodic structure*. Oxford: Oxford University Press.

Friberg, J. C., & McNamara, T. L. (2010). Evaluating reliability and validity of (Central) auditory processing tests: A preliminary investigation. *Journal of Educational Audiology*, *16*, 4-16.

Gelfand, S.A. (2001). *Essentials of audiology*. (2nd ed.) Thieme Medical Publishers.

Gelfand, S. A. (2009). *Essentials of audiology* (3rd ed.). Thieme Medical Publishers.

Gelfand, S.A. (2016). Essentials of Audiology (4th ed.). Thieme Medical Publishers.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *Qualitative Report*, 8, 597-607.

Grange, M. E. (2013*). Test retest reliability in word recognition testing in subjects with varying levels of hearing loss*. Unpublished thesis, Brigham Young University, Utah, USA.

Han, H., Lee, J., Cho, S., Kim, J., Lee, K., Choi, W. (2011). Reference sound pressure level for Korean speech audiometry. *International Journal of Audiology*, *50*, 59-62. https://doi.org/10.3109/14992027.2010.526637

Hallgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*(4) 227-237.

Hathcoat, J. D. (2013). Validity semantics in Educational and Psychological Assessment. *Practical Assessment, Research & Evaluation*, *18*,1-14.

Hanekom, T., Soer, M & Pottas, L. (2015). Comparison of the South African Spondaic and CID W-1 wordlists for measuring speech recognition threshold. *South African Journal of Communication Disorders*, *62*(1), 1-10.

Harris, R. W., Kim, E. & Egget, D. L. (2003). Psychometrically equivalent Korean bisyllabic words spoken by male and female talkers. *Korean Journal of Communication Disorders*, *8*, 217-337.

Harris, R.W., Nissen, S.L., Pola, M.G., Mcpherson, D.L., Tavartkiladze, G.A., & Egget,D.L. (2007). Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *International Journal of Audiology, 46*, 47-66.

Harris, R. W., Nielson,W. S. McPherson, D. L., Skarzynski,H & Egget, D. L. (2004). Psychometrically equivalent Polish bisyllabic words spoken by male and female talkers. *Audiofonologia, 25*, 1-15.

Health Professions Council. (2019). Guidelines for practice in a culturally and linguistically diverse South

Africa.https://www.hpcsa.co.za/Uploads/professional_boards/slh/guidelines/Guidelines for_practice_in_a_culturally_and_linguistically_divers.pdf

Honu-Mensah, C. M., Offei, Y. N., Owusu, N. A. V., & Anim-Sampong, S. (2023). Psychometrically equivalent Fante speech recognition threshold materials by a female talker. *Journal of Speech, Language, and Hearing Research, 66*(5), 1842–1852. https://doi.org/10.1044/2023_JSLHR-21-00571

Hudgins, C.V., Hawkins, J.E., Karlin, J.E., & Stevens, S.S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope, 57*, 57–89. https://doi.org/10.1288/00005537-194701000-00005

Jerger, J. (2006). Are some more equal than others? *Journal of the American Academy of Audiology, 17*, 156.

Kane, M.T. (2013).Validating the interpretations and uses of test scores. *Journal of Educational Measurement,* 50,1-73.

Kim, J., Lee, J., Lee, K. W., Bahng, J., Lee, J. H., Choi, C. H., Cho, S. J., Shin, E. Y., & Park, J. (2015). Test-Retest Reliability of Word Recognition Score Using Korean Standard Monosyllabic Word Lists for Adults as a Function of the Number of Test Words. *Journal of audiology & otology*, *19*(2), 68–73. https://doi.org/10.7874/jao.2015.19.2.68

Killman, L., Zekveld, A., Hallgren, M., & Jerker, R. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology, 5*(65), 1–9. https://doi.org/10.3389/fpsyg.2014.00651

Khoza-Shangase, K, & Mophosho, M. (2018). Language and culture in speech-language and hearing professions in South Africa: The dangers of a single story. *South African Journal of Communication Disorders*, *65*(1),7. https://doi.org/10.4102/sajcd.v65i1.594

Khoza- Shangase, K, & Mokoena, N. (2014). Speech Audiometry in Africa in Pursuit of African resource development in African languages. *I J British*, *1*(2), 1- 12

Khoza, K., Ramma, L., Mophosho, M., & Moroka, D. (2008). Digit speech reception threshold testing in Tswana/English speakers. *South African Journal of Communication Disorders*, *55*, 20-28.

Kruger, B., & Kruger, F. M. (1997). Speech audiometry in USA. In M. Martin (Ed.), *Speech Audiometry* (pp. 233-277). Whurr Publishers.

Kruger, B., & Mazor, R.M. (1987). Speech audiometry in USA. In M. Martin (Ed.), *Practical aspects of audiology: Speech audiometry* (pp. 207–235). Whurr Publishers.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*,159-174.

Lawson, G., & Peterson, M. (2011*) Speech Audiometry*. Western Michigan University. https://scholarworks.wmich.edu/books/47

Leedy, P. D., & Ormrod, J. E. (2005). *Practical research.* (8th ed.). Prentice Hall.

Liu, D., & Shi, L.F. (2013). Performance-intensity functions of Mandarin word recognition tests in noise: Test dialect and listener language effects. *American Journal of Audiology, 22*, 147–156. https://doi.org/10.1044/1059-0889(2013/12-0047)

Ma, X., McPherson, B., Ma, L. (2013). Chinese speech audiometry material: Past, present and future. *Hearing, Balance and Communication*, *11*, 52-63.

Marinova-Todd, S.H., Siu, C. K., Jensad, L.M. (2011). Speech audiometry with non-native English speakers: The use of digits and Cantonese words as stimuli. *Canadian Journal of Speech Language Pathology and Audiology, 35*, 220-227.

Martin, F. N., & Hart, D. B. (1978). Measurement of speech thresholds of Spanish-speaking children by non-Spanish speaking clinicians. *The Journal of Speech and Hearing Disorders*, *43*, 255-262.

Masalski, M., & Morawski, K. (2024). The multilingual digits-in-noise (DIN) test: Development and evaluation. *International Journal of Audiology*, *1–11*. https://doi.org/10.1080/14992027.2024.2397068

Maxwell, D.L., & Satake, E. (2006). *Research and statistical methods in communication sciences and disorders*. Singular.

MacPherson, A., & Akeroyd, M.A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing, 18*, 1–26. https://doi.org/10.1177/2331216514537722

McArdle, R., & Hnath-Chisolm, T. (2009). Speech audiometry. In J. Katz, L. Medwetsky, R. Burkard, & L. Hood (Eds.), *Handbook of clinical audiology* (6th ed., pp. 64–79). Lippincott Williams & Wilkins.

McArdle, R. H., & Wilson, R. H. (2006). Homogeneity of the 18 Quick SIN lists. *Journal of the American Academy of Audiology, 17*,157-167.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *The Journal of speech and hearing disorders*, *49*(1), 34–42. https://doi.org/10.1044/jshd.4901.34

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Medwedsky, L. (2002). Central auditory processing. In J.Katz (Ed.), Handbook of Clinical Audiology (pp. 495-524). Baltimore: Williams and Wilkins.

Melo, R. D., Menezes, D. C., Pacífico, F. A., Advíncula, K. P., & Griz, S. M. (2017). Brazilian Portuguese Hearing in Noise Test (HINT): Different interpretation criteria for individuals' responses. *Codas, 29*(1), e20160082. https://doi.org/10.1590/2317-1782/20172016082

Mendel, L.L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology, 47,* 546-553

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30*(10), 955–966. https://doi.org/10.1037/0003-066X.30.10.955

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan Publishing Co., Inc.; American Council on Education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons 'responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50* (9), 741-749.

Messick, S. Test Validity: A Matter of Consequence. *Social Indicators Research 45*, 35–44 (1998). https://doi.org/10.1023/A:1006964925094

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

MacPherson, A., & Akeroyd, M.A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing, 18*, 1–26. https://doi.org/10.1177/2331216514537722

Nissen, S.L., Harris, R.W., Jennings, L.J., Eggert, D.L., & Buck, H. (2005). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *International Journal of Audiology, 44*, 391–399. https://doi.org/10.1080/14992020500147672

Naude, A (2018). Statistical evaluation and prediction of word recognition test scores of the Foneties Verteenwoordigende Eenlettergrepige Woordlyste in Afrikaans (FVEWA*). Tydskrif vir. Geesteswetenskappe, 58* (1), 124-136. http://dx.doi.org/10.17159/2224-7912/2018/v58n1a8.

Neuman, K., Baumeister, N., Baumann, U., Sick, U., Euler, H.A., & Weibgeber, T (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *International Journal of Audiology, 51*, 157-163.

Ng, I. H.Y., Lee, Y.S., Lam, J.H.S., Van Hasslet, A. (2016). Application of Item response theory and RASCH model in speech recognition test materials. *American Journal of Audiology*, *25*(2), 142-152. doi:10.1044/2015_AJA-15-0025

Nissen, S. L., Harris, R. W., Jennings, L. J., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *International Journal of Audiology*, *44*(7), 391-399.

Nissen, S.L., Harris, R.W., & Slade, K.B. (2007). Development of speech reception materials of speakers of Taiwan Mandarin. *International Journal of Audiology*, *46*, 449-458.

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. *Journal of the Acoustical Society of America, 95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Ostergard, A. C. (1983). Factors Influencing the Validity and Reliability of Speech Audiometry. *Seminars in Hearing, 4*(3), 221-239

Panday, S. (2006). *The development of a Zulu speech reception threshold test for Zulu first language speakers in KwaZulu-Natal (KZN)* (Unpublished master's thesis). University of KwaZulu-Natal, Durban, South Africa.

Panday, S., Kathard, H., Pillay, M., & Govender, C. (2007). The development of a Zulu Speech Reception Threshold test for Zulu first language speakers in KwaZulu-Natal. *South African Journal of Communication Disorders, 54*, 111–122.

Panday, S., Kathard, H., Pillay, M., & Govender, C. (2009). The homogeneity of audibility and prosody of Zulu words for Speech Reception Threshold (SRT) testing. *South African Journal of Communication Disorders, 56*, 60–75

Panday, P., Kathard, H., Pillay, M., & Wilson, W. (2018). The internal and external consistency of a speech reception threshold test for isiZulu speakers with normal hearing sensitivity. *South African Journal of Communication Disorders 65*(1),a 556. https://doi.org/ 10.4102/sajcd.v65i1.556

Panday, S., Kathard, H., Pillay, M., & Wilson, W. (2018). First language raters' opinions when validating word recordings for a newly developed speech reception threshold test. *South African Journal of Communication Disorders 65*(1), a555. https://doi.org/10.4102/ sajcd.v65i1.555

Panday, S., Kathard, H., & Wilson, W.J. (2020). The validity of an isiZulu speech reception threshold test for use with adult isiZulu speakers. *South African Journal of Communication Disorders*, *67*(1), a690. https://doi.org/10.4102/ sajcd.v67i1.690

Paiva, C. E., Barroso, E. M., Carneseca, E. C., de Pádua Souza, C., Dos Santos, F. T., Mendoza López, R. V., & Ribeiro Paiva, S. B. (2014). A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. *BMC medical research methodology*, *14*, 8. https://doi.org/10.1186/1471-2288-14-8

Pascoe, M., Rogers, C., & Norman, V. (2013). Are we there yet? On a journey towards more contextually relevant resources in speech-language therapy and audiology. *South African Journal of Communication Disorders, 60*, 2–9. https://doi.org/10.4102/sajcd.v60i1.3

Plomp, R & Mimpen, A.M. (1979). Speech-Reception threshold for sentences as a function of age and noise level. *Journal Acoustical Society*, *66*, 1333-1342.

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: An illustration of appropriate statistical analyses. *Clinical Rehabilitation*, *12*, 187–199.

Ramkissoon, I., Proctor, A., Lansing, C., & Bilger, R.C. (2002). Digit speech recognition threshold (SRT) for non-native speakers of English. *American Journal of Audiology, 11*, 23–28. https://doi.org/10.1044/1059-0889(2002/005)

Ramkissoon, I., Estis, J. M., & Flagge, A. G. (2014). Digit Speech recognition threshold (SRT) in children with normal hearing ages (5-8 years). *American Journal of Audiology 23,* 182-189.

Shrout, P. E., & Fleiss, J. L. (1979). Intra class correlations: Uses in assessing rater-reliability. *Psychological Bulletin*, *86*, 420-428.

Riddle D.L, Finucane S.D., Rothstein J.M., & Walker M.L. (1989) Intrasession and intersession reliability of handheld dynamometer measurements taken on brain damaged patients. *Physical Therapy, 69*, 182–94.

Roup, C.M., Wiley, T.L., Safady, S.H., & Stoppenbach, D.T. (1998). Tympanometric screening norms for adults. *American Journal of Audiology*, *7*, 1059–1089. https://doi.org/10.1044/1059-0889(1998/014)

Rudwick, S. I. (2008). Shifting Norms of Linguistic and Cultural Respect: Hybrid Sociolinguistic Zulu Identities. *Nordic Journal of African Studies*, *17*(2), 23. https://doi.org/10.53228/njas.v17i2.241

Roeser, J. R., Valante, M., & Horsford-Dunn, H. (2000). *Audiology Diagnosis.* USA: Thieme Medical Publishers

Rycroft, D.K., & Ngcobo, A.B. (1979). *Say it in Zulu.* Pietermarizburg: University of Natal Press.

Scafer, E.C., Pogue, J., Milrany, T (2012). List Equivalency of the AzBio Sentence Test in Noise for Listeners with normal hearing sensitivity and cochlear implants. *Journal of American Academy of Audiology*, *23*, 501-509.

Shi, L. F., & Zaki, S. (2014). Psychometric function for NU-6 word recognition in noise: Effects of first language and dominant language. *Ear and Hearing, 35*(2), 236–245.

Silman, S., & Silverman, C.A. (1991). *Audiology diagnosis: Principles and practice*. New York: Academic Press Inc.

Rycroft, D.K., & Ngcobo, A.B. (1979). *Say it in Zulu.* Pietermarizburg: University of Natal Press.

Smits, C., Theo Goverts, S., & Festen, J. M. (2013). The digits-in-noise test: assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America*, *133*(3), 1693–1706. https://doi.org/10.1121/1.4789933

Shrout, P.E., & Fleiss, J.L. (1979). Intra class correlations: Uses in assessing rater-reliability. *Psychological Bulletin, 86*, 420-428.

Statistics South Africa. (2011). *Census 2011 statistical release – P0301.4*. Author.

Statistics South Africa. (2022). Census 2022 statistical release- - *P0160.* Author.

Streiner D.L., & Norman G.R. (1995). Health measurement scales: A practical guide to their development and use (2nd ed., pp. 104-127). Oxford: Oxford University Press.

Sreedhar, J. S., Venkatesh, L., Nagaraja, M. N., & Srinavasan, P. (2011). Development and evaluation of paired words for testing of speech recognition threshold in Telegu: A preliminary report. *Journal of Indian Speech Language & Hearing Association*, *25*(2), 128-136.

Statistics South Africa. (2018) *General household survey,* 1-203.Author

Takayanagi, S., Dirks, D.D., & Moshfegh, A. (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research, 45*, 585-597.

Thakor, H. (2020). *South African audiologists' use of speech in noise test for adults with hearing difficulties* (Unpublished master's thesis). University of the Witwatersrand, Johannesburg, South Africa

Theunissen, M., Swanepoel, D.W., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. *International Journal of Audiology, 48*(11), 747–757.https://doi.org/10.3109/14992020903082088

Van der Merwe, A., & Le Roux, M. (2014). Idiosyncratic sound systems of the South African Bantu languages: Research and clinical implications for speech-language pathologists

and audiologists. *South African Journal of Communication Disorders, 61*(1), 86 -938. https://doi.org/10.4102/sajcd.v61i1.86

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., & Giguère, C. (2005). Adaptation of the HINT (Hearing in Noise Test) for adult Canadian Francophone populations: Adaptación del HINT (Prueba de Audición en Ruido) para poblaciones de adultos canadienses francófonos. *International Journal of Audiology, 44*(6), 358–361. https://doi.org/10.1080/14992020500060875

Wang, S., Mannel, R., Newall, P., Zhang, H., Han, D (2007). Development and Evaluation of Mandarin disyllabic materials for speech audiometry in China. *International Journal of Audiology*, *46*, 719-73.

Wilson, R. H., & Carter, A. S. (2001). Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology, 12*(1), 7–14.

Wilson, W. J., & Moodley, S. (2000). Use of the CID W22 as a South African English speech discrimination test. *South African Journal of Communication Disorders, 47*(1), 57-62.

Wong, L. L., Soli, S. D., Liu, S., Han, N., & Huang, M. W. (2007). Development of the Mandarin Hearing in Noise Test (MHINT). *Ear and Hearing*, *28*(2 Suppl), 70S-74S. https://doi.org/10.1097/AUD.0b013e31803154d0

Wyatt, T. (2012). Assessment of multicultural and international clients with communication disorders. In D. E. Battle (Ed.), *Communication disorders in multicultural and international populations* (pp. 243–278). Elsevier/Mosby. http://site.ebrary.com/id/10574595

APPENDICES

Appendix A **Ethics Letter**

Appendix B **Informed Consent Document – Reliability study**

Appendix C **Information Document – Reliability study**

Appendix D **Case History Form**

Appendix E **isiZulu SRT list(Panday 2006)**

Appendix F **Rating Scale – Content Validity study**

Appendix G **Informed Consent Letter – Content Validity study**

Appendix H **Information Document – Content Validity study**

Appendix I **Information Document – Validity study – Phase 2**

Appendix J **Information Document -Validity study**

Appendix K **Declaration that permission has been granted to include Publications into the thesis.**

Appendix A. Ethics Letter

UNIVERSITY OF CAPE TOWN

Faculty of Health Sciences
Faculty of Health Sciences Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
e-mail: sumayah.ariefdien@uct.ac.za

13 December 2012

HREC REF: 652/2012

Ms S Panday
c/o A/Prof H Kathard
Hea th & Rehab Sciences
F-45
OMB

Dear Ms Panday

PROJECT TITLE: THE RELIABILITY AND VALIDITY OF A SPEECH RECEPTION THRESHOLD TEST IN QUIET FOR HEARING IMPAIRED INDIVIDUALS IN KWA ZULU NATAL

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee for review.

It is a pleasure to inform you that the HREC has **formally approved** the above mentioned study.

Approval is granted for one year till the 28 December 2013.

Please submit a progress form, using the standardised Annual Report Form, if the study continues beyond the approval period. Please submit a Standard Closure form, if the study is completed within the approval period.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC. REF in all your correspondence.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, HSF HUMAN ETHICS

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

sAriefdien

Appendix B Informed Consent Document – Reliability study

INFORMED CONSENT FORM FOR RELIABILITY STUDY

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital,
Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

TITLE OF THE PROJECT:

"The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for hearing and hearing impaired isiZulu speakers in KwaZulu-Natal"

I ____________________________________ have read (or have had read to me)

by________________) the information sheet. I understand what is required of me and I have had all my questions answered. I understand that I am only part of the first stage of this project. I do not feel that I am forced to take part in this study and I am doing so on my own free will. I know that I can change my mind to take part at any time if I so wish and nothing bad will happen to me. I also agree that if I am called back for another round of testing, I will agree to be tested again four weeks after the first test.

Signed :

____________________________	____________________
Participant	Date and place
____________________________	____________________
Researcher	Date and place.
____________________________	____________________
Audiologist	Date and place
____________________________	____________________
Witness	

Appendix C Information Document – Reliability study

APPENDIX C: INFORMATION LETTER FOR PARTICIPANTS FOR RELIABILITY AND CONVERGENT VALIDITY TESTING- PHASE 1

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital, Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

Date: ------------

Dear Participant

I am a student registered for the PhD degree in Audiology at the University of Cape Town, and I am doing a research project called "The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for normal hearing and hearing impaired isiZulu speakers in KwaZulu-Natal" My supervisor is Prof. H Kathard from the University of Cape Town's Division of Communication Sciences and Disorders, Dr. Wayne Wilson from the University of Queensland (Australia) and Dr Mershen Pillay.

The main reason for doing this study is that at the moment, there is no good or accurate isiZulu speech test in KwaZulu-Natal to find out how isiZulu speakers hear speech or words. I would like to invite you to take part in this study because you are an isiZulu speaker from KwaZulu-Natal and are between the age of 18-60 years. If you agree to take part in this study, you will be part of the first part of this project and you will have a free hearing test. You will answer some questions about your general health, have your ears checked for any blockages, listen to tones and English and isiZulu words to see how well you hear. All of these tests are the same that you would have at the hospital or clinic when you visit an audiologist. These tests will not cause you any pain, harm or be uncomfortable. We have received permission from the Faculty of Health Sciences Human Research and Ethics committee of the University of Cape Town to do this study.

If you agree to take part in this study, you will be given an appointment go to the audiology clinic at the University of KwaZulu-Natal on the Westville Campus in Durban for approximately one hour. You may be asked to come back for a second time, four weeks after the first appointment for another 1 hour. If you agree to come back for a second time you will be given an appointment card with the date for the next appointment. You will be given all information in isiZulu as well as a letter about what you have to do in the study. You will be given an explanation of your hearing test results by a trained audiologist (a hearing specialist) who speaks isiZulu. You will be sent to the public hospital or clinic closest to where you live or if you decide, to a private hospital or practitioner, if you need more testing or a need to see another medical specialist, but you will have to pay for this.

All your results and any information we have about you will be kept private and not given to anyone. We will also not use your name in any of our work, and will add the results of all the people together in the study, as this study will be done on many people. Instead of using your name when we look at your results, we will use a code, so that we never know whose information or test results we are looking at. All the information we get from you will only be used for this study, and when we write the results up for other hearing specialists and professionals to read and when we speak about the results at conferences so that other people can hear or read about what we have done. All information that we obtain from your tests will be kept in the researcher's office in a locked cabinet and will remain there for at least 5 years after the end of the study.

You will be given money (the price of taxi fair or bus fair) for your travel to and from the clinic at the University in Westville. We will make sure that you are completely comfortable at all times during your time in the study. You do not have to take part and you can stop at any time if you do not want to carry on any more and any services that you receive at your normal clinic will not be affected. If you need more information you can contact Ms. S. Panday on the 031-2607623 or pandayse@ukzn.ac.za or the project supervisor Prof. Harsha Kathard on 021- 4066593 harsha.kathard@uct.ac.za. You can also contact the Faculty of Health Sciences Human Research Ethics Committee of University of Cape Town on 021-4066346 or fax 021-4478955 or the postal address is Faculty of Health Sciences, Private Bag, Observatory 7935, if you have any further queries or concerns about your participation on this study.

Your help with this project will be appreciated. We hope you will think about joining the study. If you do decide to join, please sign the attached form which gives us permission to include you in this study.

Yours sincerely,

____________________	________________________	__________
Ms S. Panday (PhD student)	A/Prof. H Kathard (Supervisor)	Dr. W. Wilson (Co supervisor)

Appendix D. Case History Form

CASE HISTORY QUESTIONNAIRE

SECTION A

BIOGRAPHICAL DATA

1. PARTICIPANT NUMBER: _______________________________
2. DATE OF BIRTH: __________________________________
3. AGE: ____
4. GENDER: ____
5. FIRST LANGUAGE __________________________________
6. RESIDENTIAL ADDRESS _____________________________

____

SECTION B

GENERAL MEDICAL AND NEUROLOGICAL HISTORY

1. Do you or have you suffered from any medical problems? For example Rubella, meningitis etc .

   Yes ☐ No ☐

   If yes, describe when and for how long you had the condition.

____
____
____

2. Are you or have you previously received treatment for any of the above?

   Yes ☐ No ☐

   If yes, state the nature and duration of the treatment.

____
____
____

4. If you are taking or have taken any form of medication, please state:

a. The name of the drugs taken __________________________________

b. Dosage taken: 5mg 5ml OR 10mg 10ml OR > 10mg > 10ml

c. Frequency of consumption: 1 2 3 4 >4 times a day

d. Duration of treatment: <1 2 >3 months

5. Have you been hospitalised or received treatment for a prolonged period for any medical, surgical or neurological problems?

Yes ☐ No ☐

If yes, please describe.

___

___

6. Do you suffer with headaches?

Yes ☐ No ☐

If yes, answer the following:

a. Type: Migraine Tension Other ____________________________

b. Onset: Morning Midday Afternoon Night

c. Duration: < 1 hour > 1 hour

d. Frequency: 1 2 3 per day OR continual

e. Severity: Mild Moderate Severe Very Severe

7. Do you or have you previously suffered from any dizzy spells?

Yes ☐ No ☐

If yes, answer the following:

a. Onset: Morning Midday Afternoon Night

b. Duration: < 1 min > 1 min

c. Nature: When experiencing dizziness, do you tend to fall to the
Right Left Forward Backward

d. Do you feel that objects around you tend to spin?

Yes ☐ ☐

e. Do you feel that you are spinning around?

Yes ☐ ☐

f. Frequency: 1 2 3 per day OR continuously

g. Severity: Mild Moderate Severe Very Severe

SECTION C

AUDIOLOGICAL HISTORY

1. Do you experience difficulty in hearing?

Yes ☐ No ☐

If yes, describe your difficulties:

____

____

____

____

____

2. Have you experienced any pain in your ears?

Yes ☐ No ☐

If yes, what diagnosis was given upon visiting a doctor?

____

____

____

3. Do you have any discharge from your ears?

Yes ☐ No ☐

4. Did you suffer from any other nose, ear or throat problems?

Yes ☐ No ☐

If yes, describe your problem:

____

____

____

5. Do you experience difficulty listening in background noise?

Yes ☐ No ☐

If yes, describe your problem:

____

____

6. Do you hear any unusual sounds or noises in your ears?

Yes ☐ No ☐

If yes, answer the following:

a. Ear: Left Right Both

b. Nature: High pitched Low pitched Ringing Pulsating Roaring

c. Onset: Morning Midday Afternoon Night

d. Frequency: Intermittent Continuous

e. Severity: Soft Loud Very Loud

f. Does this worry you?

Yes ☐ No ☐

7. Do you live in a noisy environment?

Yes ☐ No ☐

If yes, state the source and type of noise that you are exposed to.

____

____

Does this noise annoy or disturb you?

Yes ☐ No ☐

Please describe.

____

____

8. Do you participate in any hobbies or sports that involve exposure to very loud sounds example shooting?

Yes ☐ No ☐

If yes, please describe.

____

____

9. Have you experienced head or ear injury at any time?

Yes ☐ No ☐

If yes, please describe.

____
____
____

10. Are you particularly sensitive to loud noises; do they annoy, irritate or cause you any discomfort?

Yes ☐ No ☐

If yes, please elaborate.

____
____
____

11. Do you have difficulty listening to male voices?

Yes ☐ No ☐

12. Do you have difficulty listening to female voices?

☐

SECTION D

FAMILY HISTORY

1. Does anyone in your family have a hearing problem?

Yes ☐ No ☐

Please describe.

____
____
____

3 **SECTION E**

4 OCCUPATIONAL HISTORY

1. Please describe your current and previous working environments?

____
____
____
____

2. How long have you worked in each environment?

____
____
____

3. Please state any exposure to excessive noise in each environment.

____
____
____

I wish to inform you that the information you have supplied will be screened and selected individuals will be required to undergo a complete audiological evaluation. As you have already indicated your willingness to participate in such a program, we will be contacting you at the given address/telephone number.

THANK YOU FOR YOUR TIME AND WILLINGNESS TO PARTICIPATE IN OUR RESEARCH PROJECT.

-------------------------------	-----------------------------------
Ms S. Panday	A/Prof. H. Kathard
Researcher (PhD Student)	Supervisor
	(University of Cape Town)

Appendix E isiZulu SRT list

<u>ISIZULU SRT WORD LIST</u>
(Panday 2006)

Banga
Gxeka
Cinga
Faka
Thela
Linda
Khaba
kheta
Thatha
Donsa
Washa
Chela
Xola
Yonga
Yona
Veza
Wina
Khanya
Shada
Geza
Khipa
Thola
Jeza
Qonda
Thenga
Loya
Minya
Yeka

	PTA	*SRT*
Right Ear		
Left Ear		

Appendix F Rating Scale – Content Validity study

Instructions to participants

Please listen to the recording of the following isiZulu word list. You will listen to the recording three times before you fill out the rating scale. You are required to rate this recording for the following aspects on a scale of 1 to 5 where 1 = strongly agree, 2 agree, 3 neutral, 4 disagree, 5 strongly disagree. Please mark off with an X for each category.

Words	**Pitch/Tone** (sounds like a low pitch verb in isiZulu)					**Clarity/articulation** (The recording of the word is clear and easily understood)					**Naturalness/Dialect** (The recording of the word sounds natural and is similar to how isiZulu is spoken every day in KZN)					**Speech rate** The speakers rate of speech is natural/normal					**Quality** The quality of the recording is good.				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Banga																									
Gxeka																									
Cinga																									
Faka																									
Thela																									
Linda																									
Khaba																									
Kheta																									
Thatha																									
Donsa																									
Washa																									
Chela																									
Xola																									
Yonga																									
Yona																									
Veza																									
Thenga																									
Wina																									

Words	**Pitch/Tone** (sounds like a low pitch verb in isiZulu)					**Clarity/articulation** (The recording of the word is clear and easily understood**)**					**Naturalness/Dialect** **(**The recording of the word sounds natural and is similar to how isiZulu is spoken every day in KZN)					**Speech rate** The speakers rate of speech is natural/normal					**Quality** The quality of the recording is good.				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Khanya																									
Shada																									
Geza																									
Khipha																									
Thola																									
Jeza																									
Qonda																									
Loya																									
Minya																									
Yeka**																									

Appendix G Informed Consent Letter – Content Validity study

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital,
Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

TITLE OF THE PROJECT:

"The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for hearing and hearing impaired isiZulu speakers in KwaZulu-Natal"

I ____ have read (or have had read to me) by_________________ the information sheet. I understand what is required of me and I have had all my questions answered. I understand that I am only part of the first stage of this project. I do not feel that I am forced to take part in this study and I am doing so on my own free will. I know that I can change my mind to take part at any time if I so wish and nothing bad will happen to me.

Signed :

________________________________ ____________________
Participant Date and place

________________________________ ____________________
Researcher Date and place.

________________________________ ____________________
Audiologist Date and place

________________________________ ____________________
Witness Date and place

Appendix H Information Document – Content Validity study Phase 1

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital, Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

Date: ------------

Dear Participant

I am a student registered for the PhD degree in Audiology at the University of Cape Town, and I am conducting a research project called "The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for normal hearing and hearing impaired isiZulu speakers in KwaZulu-Natal" My supervisor is Prof. H Kathard from the University of Cape Town's Division of Communication Sciences and Disorders, and Dr. Wayne Wilson from the University of Queensland (Australia) and Dr Mershen Pillay.

The main reason for doing this study is that at the moment, there is no good or accurate isiZulu speech test in Kwa Zulu-Natal to find out how isiZulu speakers hear speech or words. I would like to invite you to take part in this study because you are an isiZulu speaker from Kwa Zulu-Natal and are between the age of 18-60 years. If you agree to take part in this study, you will be part of the first part of this project and you will have a free hearing test. You will answer some questions about your general health, have your ears checked for any blockages, listen to tones to see how well you hear You will also be asked to listen to a recording of isiZulu words and you will have to indicate on a form if these words sound the same as the way isiZulu is spoken in Kwa Zulu-Natal. We received permission from the Faculty of Health Sciences Human Research Ethics committee of the University of Cape Town to do this study.

If you agree to take part in this study, you will go to the audiology clinic at the University of KwaZulu-Natal on the Westville Campus in Durban for approximately 1 hour. You will be given an appointment card. You will be given information in isiZulu about what to do in this study. You will be given an explanation of your hearing test results by a trained audiologist (a hearing specialist). You will be sent to the public hospital or clinic closest to where you live or if you decide, to a private hospital or practitioner, if you need more testing or a need to see another medical specialist,, but you will have to pay for this on your own.

All your results and any information we have about you will be kept private and not given to anyone. We will also not use your name in any of our work, and will add the results of all the people together in the study, as this study will be done on many people. Instead of using your name when we look at your results, we will use a code, so that we never know whose data we

are looking at. All the information we get from you will only be used for this study, and when we write the results up for other hearing specialists and professionals to read and when we speak about the results at conferences so that other people can hear or read about what we have done. All information that we obtain from your tests will be kept in the researcher's office in a locked cabinet and will remain there for at least 5 years after the end of the study.

Appendix I Informed Consent – Validity study – Phase 2

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital,
Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

TITLE OF THE PROJECT:

"The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for hearing and hearing impaired isiZulu speakers in KwaZulu-Natal"

I ____ have read (or have had read to me) by_________________ the information sheet. I understand what is required of me and I have had all my questions answered. I understand that I am only part of the second stage of this project. I do not feel that I am forced to take part in this study and I am doing so on my own free will. I know that I can change my mind to take part at any time if I so wish and nothing bad will happen to me.

Signed :

________________________________ ____________________
Participant Date and place

________________________________ ____________________
Researcher Date and place.

________________________________ ____________________
Audiologist Date and place

________________________________ ____________________
Witness Date and place

Appendix J –Information Document -Validity study

Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Divisions of Communications Sciences and Disorders, Nursing and Midwifery, Occupational Therapy, Physiotherapy

F45 Old Main Building, Groote Schuur Hospital, Observatory 7925
Tel: +27 (0) 21 406 6401 Fax: +27 (0) 21 406 6323
Internet: www.uct.ac.za

Date: ------------

Dear Participant

I am a student registered for the PhD degree in Audiology at the University of Cape Town, and I am conducting a research project called "The reliability and validity of a Speech Reception Threshold (SRT) test in quiet for normal hearing and hearing impaired isiZulu speakers in KwaZulu-Natal" My supervisor is Prof. H Kathard from the University of Cape Town's Division of Communication Sciences and Disorders, and Dr. Wayne Wilson from the University of Queensland (Australia) and Dr Mershen Pillay.

The main reason for doing this study is that at the moment, there is no good or accurate isiZulu speech test in KwaZulu-Natal to find out how isiZulu speakers hear speech or words. I would like to invite you to participate in this study because you are a first language isiZulu speaker from KZN and are between the age of 18-60 years and have a hearing loss . If you agree to take part in this study, you will form part of the second part of this project and will have a free hearing test to check your hearing levels again. This would involve filling out a questionnaire, checking your ears for any blockages and listening to tones and English and isiZulu words. All of these procedures are the same that you would have at the hospital or clinic when you visit an audiologist and will not cause you any pain, harm or discomfort.

If you agree to take part in this study, you will go to the audiology clinic at the University of KwaZulu-Natal on the Westville Campus in Durban. These tests will take about 2hours of your time. You will be given short breaks between the tests so that you can have rest. You will be given information in your first language as well as a letter about what you have to do in this study. You will be given an explanation of your hearing test results by a trained audiologist (a hearing specialist).

All your results and any information we have about you will be kept private and not given to anyone. We will also not use your name in any of our work, and will add the results of all the people together in the study, as this study will be done on many people. Instead of using your name when we look at your results, we will use a code, so that we never know whose data we are looking at. All the information or test results we get from you will only be used for this study All the information we get from you will only be used for this study, and when we write the results up for other hearing specialists and professionals to read and when we speak about the results at conferences so that other people can hear or read about what we have done. All

information that we obtain from your tests will be kept in the researcher's office in a locked cabinet and will remain there for at least 5 years after the end of the study.

You will be given money (the price of taxi fair or bus fair) for your travel to and from the clinic at the University in Westville. We will make sure that you are completely comfortable at all times during your time in the study. You do not have to take part and you can stop at any time if do not want to carry on any more and any services that you receive at your normal clinic will not be affected. If you need more information you can contact Ms. S. Panday on the 031-2607623 or pandayse@ukzn.ac.za or the project supervisor Prof. Harsha Kathard on 021-4066593 or harsha.kathard@uct.ac.za. You can also contact the Faculty of Health Sciences Human Research Ethics Committee of University of Cape Town on 021-4066346 or fax 021-4478955 or the postal address is Faculty of Health Sciences, Private Bag, Observatory 7935, if you have any further queries or concerns about your participation on this study.
Your help with this project will be much appreciated. We hope you will think about joining the study. If you do decide to join, please sign the attached consent form.

Yours sincerely,

____________________	________________________	___________
Ms S. Panday	A/Prof. H Kathard	Dr.W.Wilson
(PhD student)	(Supervisor)	(Co-supervisor)

Appendix K Declaration that permission has been granted to include Publications into the thesis.

"I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s)"

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

a. Panday, S., Kathard, H., Pillay, M., & Wilson, W. (2018). The internal and external consistency of a speech reception threshold test for isiZulu speakers with normal hearing sensitivity. *South African Journal of Communication Disorders* 65(1),a 556. https://doi.org/ 10.4102/sajcd.v65i1.556
b. Panday, S., Kathard, H., Pillay, M., & Wilson, W. (2018). First language raters' opinions when validating word recordings for a newly developed speech reception threshold test. *South African Journal of Communication Disorders* 65(1), a555. https://doi.org/10.4102/ sajcd.v65i1.555
c. Panday, S., Kathard, H., & Wilson, W.J. (2020). The validity of an isiZulu speech reception threshold test for use with adult isiZulu speakers. *South African Journal of Communication Disorders,* 67(1), a690. https://doi.org/10.4102/ sajcd.v67i1.690

SIGNATURE: DATE:22/12/2024

STUDENT NAME: Seema Panday

STUDENT NUMBER: PNDSEE001

Appendix L isiZulu SRT word list for KwaZulu-Natal (Panday 2025)

banga
gxeka
faka
thela
linda
khaba
kheta
thatha
donsa
washa
chela
xola
yonga
yona
veza
wina
khanya
shada
geza
khipa
thola
jeza
qonda
thenga
loya
minya
yeka

	PTA	*SRT*
Right Ear		
Left Ear		