Assessment of Revised Patient Reported Outcome Measures for Improved Readability

Roy de Hamel

Thesis submitted in partial fulfilment of the requirements for the Degree

of Master of Audiology

Department of Communication Disorders University of Canterbury

2019

Acknowledgements

I would first like to sincerely thank Dr Rebecca Kelly-Campbell for being a better, more supportive and helpful supervisor than I could have possibly hoped for. I have so appreciated your insight, knowledge and navigation from start to finish of this project. You pushed me when I needed pushing, and gave me space when I needed it most. You are the absolute best.

I would also like to thank Dr Kim Wise for helping edge the door to audiology open to me, I can quite honestly say that your help changed the course of my life, and I will remain thankful for that throughout my career.

I would like to thank the entire Maud faculty staff for actively supporting me through a tough season of personal bereavement while writing this thesis. Your understanding and words of encouragement meant more than you may know, and were far beyond your call of duty. The subject librarians also owed my highest praise for their consistently incredible help.

I would like to thank and congratulate the Maud class of 2019, it was a genuine pleasure getting to study alongside you all over these two years, you have become my great friends, and I could not think of a better, or more generally enjoyable cohort to do life with.

Penultimately, I would like to thank my family for being ever loving and supportive throughout the many years I have now spent at university, for constantly critiquing my use of the English language in my youth, and for not acting too surprised when I submitted my thesis on time.

Finally, thank you to my darling wife, Sophie, this thesis is dedicated to you. You are the best human I have ever met, and you know at least as well as me that I would not have even attempted, let alone survived these last two years without you. You are my everything.

Abstract

Purpose

The purpose of this study was to revise the wording of patient-reported outcome measures (PROMs) until they fell within the recommended readability levels for health-related materials. The subsequent scores of the revised PROMs were then to be compared to the original versions of the PROMs in order to assess the effect of the revisions. The revised scores were also to be compared with a second copy of the revised version in order to assess the consistency of the results.

Method

Three published PROMs were selected and were revised to within the recommended readability levels for health-related materials. The PROMs were revised until they met the recommended 6th reading grade level (RGL) according to the F-K readability formula. Three copies of the PROMs were distributed to participants in the United States of America in a random order (two revised versions and one original) and the results were compared.

Results

The results revealed that there was a much larger difference between the original and revised versions than there was between the two copies of the revised versions. However, the results also revealed that the study was overpowered, preventing the results from being statistically conclusive.

Conclusion

The results suggest that if PROMs are revised to within the recommended levels for health-related materials, then clients will give more clinically valid, consistent answers than what they do with the current versions. However, further research is required in order to corroborate these results.

List of Abbreviations

AC	Air Conduction
AD	Adjustment
AH	Advanced Handling
AL	Aided Listening
ANCOVA	Analysis of Covariance
ANOVA	Analyses of Variance
APD	Auditory Processing Disorder
APHAB	Abbreviated Profile of Hearing Aid Benefit
ART	Acoustic Reflex Threshold
BC	Bone Conduction
BEPTA	Better Ear Puretone Average
BH	Basic Handling
BKB-SIN	Bamford-Kowal-Bench Speech-in-Noise Test
dB HL	Decibels Hearing Level
DOSO	Device Oriented Subjective Outcome Scale
F-K	Flesch-Kincaid Grade Level
HA	Hearing Aid
HA-SE	Hearing Aid Self-Efficacy
HHI	Hearing Health Indicator
HHIE	Hearing Handicap Inventory for the Elderly
HI	Hearing Impairment
HINT	Hearing in Noise Test
ICF	International Classification of Functioning, Disability and Health
IEEE	Institute of Electrical and Electronics Engineers
IOI-HA	International Outcome Inventory for Hearing Aids

MARS-HA	Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids
ME	Middle Ear
OAEs	Otoacoustic Emissions
Oth	Impact on Others
PROMs	Patient-Reported Outcome Measures
PTA	Puretone Audiometry
QoL	Quality of Life.
QuickSIN	Quick Speech in Noise Test
RAL	Residual Activity Limitations
REAG	Real-Ear Aided Gain
REM	Real-Ear Measurement
RGL	Reading Grade Level
RM- MANOVA	Repeated Measure Multivariate Analysis of Variance
RPR	Residual Participation Restrictions
SADL	Satisfaction with Amplification in Daily Life
SAHL	Short Assessment of Health Literacy
Sat	Satisfaction
SE	Self-Efficacy
SMOG	Simple Measure of Gobbledygook
SPSS	Statistical Package for Social Sciences
WEPTA	Worse Ear Puretone Average
WHO	World Health Organization
WIN	Words in Noise

Contents

List of figures

List of Tables

Chapter 1 : Introduction and Review of Literature

1.1 Overview

The field of clinical audiology is becoming increasingly patient-oriented (Cox, 2005; Gilligan & Weinstein, 2014), with steps being taken to include the patient in their own treatment. As such, Patient Reported Outcome Measures (PROMs) are a regular feature in audiology clinics. However, due to the nature of self-reporting, a patient's personal health literacy and comprehension act as mediators to clinical outcomes (Baker, 2006). Therefore, the validity of any given PROM is inextricably linked to the health literacy of the individual completing it.

In an effort to mitigate this effect, recommended readability levels for health-related information have been established, though previous studies indicate that the majority of PROMs fall outside this range (Atcherson, Richburg, Zraick, & George, 2013; Atcherson, Zraick, & Brasseux, 2011; Douglas & Kelly-Campbell, 2018; Rebecca J Kelly-Campbell, Atcherson, Zimmerman, & Zraick, 2012).

In light of this information, it can therefore be supposed that reducing the health literacy level required to comprehend PROMs can help improve their clinical validity and consistency.

1.2 Hearing Impairment

The term "hearing impairment" (HI) is incredibly broad when isolated from context or degree, and is not merely a dichotomous distinction between being impaired or not (Dahl, 2002). From an audiological or medical perspective, the origin of the impairment may be conductive; a mechanical issue within the outer and/or middle ear (Katz, 2014), and therefore possibly be treatable with medical or surgical intervention. Alternatively, the issue could be

permanent, arising from the inner ear itself, or even originating further up the neural pathway to the brain, known as sensorineural impairment (Katz, 2014).

However, more importantly (at least to this study), any one of these examples may have a profoundly varied impact on an individual's interaction with their sociocultural surroundings, and quality of life. As such, sub-definitions for the term "hearing impairment" are necessitated.

1.2.1 Definitions

There are three commonly used scales of HI, which are calculated from the average puretone thresholds at 500, 1000 and 2000 Hz (Schlauch & Nelson, 2014). These are the Goodman (1965) scale, the Northern and Downs (2002) scale and the Jerger and Jerger (1980) scale. Although there is variation between the references as to the upper limit of normal hearing, the World Health Organization (WHO) recognises the more liberal definition of Goodman (1965), of 25 dB HL or better (WHO, 2018). This, therefore, defines hearing impairment as having at least one ear with thresholds outside that range. A disabling HI is clarified by WHO as any loss greater than 40 dB HL in the better ear in adults. For the purpose of this study this will function as the working definition. It is important to note, however, that a true definition of hearing impairment is not static and should be considered within a societal context (Dahl, 2002), therefore dictating that the definition will be influenced by factors pertaining to that context.

1.2.2 Prevalence

It is estimated by WHO (2018) that there are currently 466 million people living with a disabling HI globally. The study of Stevens et al. (2013) (based on an average of 35 dB impairment or worse in the better hearing ear) evaluated (at a 95% confidence interval) that

prevalence in adults (>15 years of age) was between 7.7 – 12.2% in females, and 9.7 – 16.2% in males. The same study showed positive correlation between HI and the male sex, middle and low-income regions, and age.

1.2.3 Management

The management of HI is primarily delivered through aural rehabilitation (AR), though it is not limited to this exclusively. Boothroyd (2007, p. 63) holistically defines AR as "the reduction of hearing-loss-induced deficits of function, activity, participation, and quality of life through a combination of sensory management, instruction, perceptual training, and counselling." In the context of audiology, sensory management via hearing aids (HAs) or other assistive listening devices is often the primary focus. Without the inclusion of instruction, perceptual training and/or counselling, however, an optimal result cannot be assumed (Boothroyd, 2007).

1.3 ICF Model

The practice of classifying people according to their symptoms and functional ability is not a new concept, and can be useful in terms of providing systematic information to health professionals, and selecting helpful interventions (Dahl, 2002). However, while noting that, Dahl (2002) recommends that classifying impairments, and those living with them, in the same way we classify diseases is nonsensical. In this way, the International Classification of Functioning, Disability and Health (ICF) presents a helpful framework with which to approach the classification process. This process takes into account a broader, multifactorial understanding of human functioning, merging the biomedical paradigms of the physical or mental impairment, with the social paradigms of an individual living in society (Danermark et al., 2010). The ICF states its intention as being a multipurpose tool, serving multiple disciplines within various sectors, aiding classification and conceptualisation of impairment

and health related states (WHO, 2001). There are several key aspects of conceptualisation and categorisation defined in the ICF that are particularly applicable to audiology.

1.3.1 Impairment

The ICF defines impairment as "a loss or abnormality in body structure or physiological function (including mental functions)" (WHO, 2001). This definition utilises the term 'abnormality' in strict reference to a distinct discrepancy between said abnormality and an established statistical norm, such as 'normal hearing' in an audiological context. The ICF emphasises that the term should only be used in such a sense. It is worth further considering, that though an individual's impairment is closely linked to the associated health condition, it is not simply a consequence of it (Danermark et al., 2010; Stucki & Grimby, 2004), as there are multiple factors (described later) beyond just that of the physical nature of the condition that influence the severity of the associated impairment.

1.3.2 Activity Limitations and Participation Restrictions

The ICF takes the term disability, that was used in WHO's 1980 publication; International Classification of Disabilities and Handicaps (ICIDH), and updates the phrasing to "activity limitation" (WHO, 2001). The ICF defines any issue an individual may have in accomplishing a task or activity due to their impairment as an activity limitation. In the case of audiology, the disability generally pertains to the loss of hearing to a debilitating extent. The scope of the limitation can be anything, such as hearing, and can range from a mild to severe deviation from what would be expected in terms of quality or quantity if the activity were performed by someone without an impairment (WHO, 2001).

When an individual's impairment impedes their ability to involve themselves in various life activities within the context of culture or society (when compared to someone

without an impairment), the ICF defines this as a participation restriction. This term is updated from the 1980 ICIDH definition "handicap" (WHO, 2001).

1.3.3 Environmental Barriers & Facilitators

An individual's experience in life is not isolated to their existence alone, but is profoundly affected by their interactions with others and with their given environment. Stucki et al. (2002) suggest that activity and participation restrictions must refer to "the environmentally adjusted inherent ability of an individual", as a person cannot remove themselves from the context of their environment. In the case of impairment, the environment of an individual can have a positive or negative influence. In the case of the ICF, environmental factors refer to the external and extrinsic context within which an individual lives. This context will have a significant effect on the functional ability of an individual. These factors can be physical, social, societal, political, legal, cultural, technological, or any other external factor pertaining to the functional capabilities of a person (WHO, 2001).

When these factors are positively influential (i.e. improve functionality or reduce the limitation of an impairment) and cause an individual's capacity or capability to be greater than what it would otherwise be, then it is considered to be an environmental facilitator. These facilitations can either be via the presence a helpful factor, such as a technology, or the absence of a negative factor, such as a social stigma, in that environment (WHO, 2001).

When a factor inhibits a person's capacity or ability to function due to its presence or absence in their environment, then it is referred to as an environmental barrier. These factors function to exacerbate a disability beyond what it would normally be. Examples of these could include physical barriers, such as physical inaccessibility, or societal barriers in which social attitudes towards a disability, lack of relevant services, systems and policies cause an individual's performance limitations to be greater than they would otherwise be (WHO,

2001). In the case of hearing impairment, there are both physical barriers, in terms of the malfunctioning of the hearing system itself, and also societal barriers, in that there are many instances in modern society that do not cater for the hearing impaired, and consequently exclude those who are.

1.3.4 Third-Party Disability

When an individual's limitations (activity, participation, environmental or other) begin to impose restriction on those around them, then it is defined as third-party disability. Those affected are not limited due to their own bodily, physiological or mental dysfunction, but by virtue of their close involvement in the life of the individual who is, in fact, impaired (WHO, 2001). An example of this would be when a spouse with normal hearing ceases to attend social events, because their partner with hearing impairment feels too embarrassed to attend, and chooses not to accompany them. The World Health Organization intends to further develop the application of the ICF to the area of third-party disability.

1.4 Audiological Assessment

The field of audiology is no longer merely focussed on the identification of hearing impairment, as it was in its infancy (Katz, 2014), but is now focused on overcoming impairment and increasing quality of life for patients. It is difficult to do this, however, without first ascertaining the issue within the hearing system. As such, diagnostic audiology uses audiologic testing to determine and locate the issue within the auditory system, and even to potentially provide further insight into the disorder itself (Kreisman, Smart, & John, 2014). The intention of this process is to help mitigate the effect of the disorder for the person, and so benefit their quality of life. As such, the audiological test battery functions to assess the extent of impairment across the previously mentioned ICF classification categories. The test battery should be compiled of subjective behavioural and self-reported tests, as well as

objective measures in order to ensure the most accurate diagnosis. Disturbances in the auditory structures and systems are measured by psychoacoustic testing, such as puretone audiometry or speech testing, or else via physioacoustic tests such as immittance testing and otoacoustic emissions (Engdahl, Tambs, & Hoffman, 2013). These tests should be used in conjunction with one another in order to cross-check the results, as increasing the number of cross-checks simultaneously decreases the likelihood of an incorrect diagnosis, while increasing the chance of an accurate identification of a disorder (Kreisman et al., 2014).

1.4.1 Impairment

The 'gold standard' of audiometric testing, according to Sindhusake et al. (2001), is puretone audiometry (PTA). PTA is a behavioural test designed to assess the presence and extent of a HI, and is administered via the air conduction (AC), or by bone conduction (BC) pathways. AC testing is performed using either insert, supra aural, or circumaural headphones in order to assess the functionality of outer and middle ear sound conduction. BC testing utilises a small vibrator placed on the mastoid process or forehead to stimulate the cochlea directly, and so tests for sensorineural HI as separate from conductive HI (Schlauch & Nelson, 2014).

Otoacoustic emissions (OAEs) provide an objective view into the mechanism of outer hair cell function within the cochlea (Kreisman et al., 2014; Robinette & Glattke, 2007). As opposed to PTA, OAEs are not behavioural, and so can provide independent information about cochlea health. They are also beneficial in that they are not affected by issues such as background noise, lack of motivation, attention loss, or other psychological factors which may affect the results of a behavioural test (Engdahl et al., 2013). OAEs are a useful crosscheck to PTA, and can further help to diagnose conductive components in HI, as well as

non-organic HI, auditory neuropathy spectrum disorder, and retrocochlear pathologies (Kreisman et al., 2014).

Further objective testing can be achieved with the use of tympanometry. Tympanometry allows us to assess what we cannot see via visual inspection or otoscopy alone (Welling & Ukstins, 2015). The information given to us by tympanometry offers a window into the health of the middle ear (ME) system behind (and including) the tympanic membrane. This is achieved by testing the ME's dynamic reaction to changes in atmospheric pressure (Hunter & Sanford, 2014). Like OAEs, this assessment is useful as a crosscheck to PTA, as it does not require participation or attention from the client, and so provides objective information (Welling & Ukstins, 2015).

1.4.2 Functional Consequences of Hearing Impairment

The functional consequence of HI is activity limitation (WHO, 2001), therefore meaning the limitation is defined by an individual's interactions with their environment. PTA may provide a more objective analysis of the auditory system, but is isolated from a person's environment and so is limited in its ability to evaluate activity limitation or participation restriction. As such, assessing these limitations often requires more than simple psychoacoustic testing (Engdahl et al., 2013). Some examples of tests designed to measure activity limitation are speech in quiet testing, speech in noise testing, and self-assessment. These tests are designed to measure the *effect* of HI, not merely the presence of it (Welling & Ukstins, 2015).

Speech in quiet testing is regularly used in audiological assessment, commonly using phonemically balanced word lists as test stimuli (Welling & Ukstins, 2015). This method is used to attempt to quantify someone's speech recognition ability, assessing both speech sensitivity and speech clarity (McArdle & Hnath-Chisolm, 2014). Assessing the clarity of

speech is imperative, as HI can also introduce distortional elements to the auditory system (Welling & Ukstins, 2015), and because speech understanding is dependent on cognition as much as it is on sound perception alone.

Speech in noise testing is useful in that adults with HI commonly express frustration at struggling to understand a talker when in the presence of background noise. Though speech in quiet tests are useful in evaluating cognition and comprehension of speech, they are not good predictors of word recognition ability for a person with HI in the presence of background noise (McArdle & Hnath-Chisolm, 2014). McArdle and Hnath-Chisolm (2014) note that people with sensorineural hearing impairment generally require a signal that is 10-12 dB louder than the competing noise in order to achieve a 50% score in word recognition tests. For normal hearing listeners, only a 2-6 dB differential between signal and noise is necessitated to achieve the same score. As such, assessing listening performance in noise can be valuable in establishing the required signal to noise ratio for an individual to understand speech.

Self-reporting can be a beneficial tool for assessing activity limitation due to HI. Even though there are correlations between PTA average and self-assessed activity restriction (Engdahl et al., 2013), the standard audiological test battery cannot function as a true measure of real-world impairment. Kreisman et al. (2014) assert that though the role of a test battery is imperative, it can never provide a complete picture to any audiologist when used on its own, and as such, self-assessment should be used as a cross-check for the test battery itself. Noble (2013), further notes that a distinct advantage that self-reporting has over more objective measures, is that it can underline inconsistencies within the results of traditional test batteries, and therefore lead to a more accurate diagnosis. For example if a person presents with normal thresholds on their audiogram, but still reports HI, then that client could now be assessed for auditory processing disorder (APD) and treated as such (Noble, 2013). Without self-reported,

functional cross-checks, such a diagnosis would not eventuate. Self-reports are simple to administer, while also being inexpensive and quick to complete (Sindhusake et al., 2001), and in that audiologists are often pressed for time in clinic, this is of significant benefit (Ingo et al., 2017).

By virtue of self-assessment being able to evaluate the effect of HI holistically, it is also the primary tool for evaluating participation restrictions, environmental barriers and facilitators, and third-party disability. In fact, if the objective is to assess the disability and handicap, then Noble (2013) recommends that self-reporting is essentially obligatory. This places an onus on the validation of self-reported measures, as there are several potentially confounding factors, such as the individual's mental health state, preconceived ideas about HI, and health literacy skills (Engdahl et al., 2013). Other obvious issues include the fact that in self-assessment a person can make themselves appear more or less impaired than they really are (as opposed to in behavioural evaluation wherein they can only feign being worse than reality) (Noble, 2013). Sindhusake et al. (2001) therefore emphasise that though self-assessment is important in the validation of the audiological test battery itself, the standard audiometric measures are equally as important in order to authenticate self-reported information. Therefore self-assessment should not be used in isolation, but as part of a test battery as a cross check to psychoacoustic and objective information. This is to say; the general principle in all forms of assessment is that the values obtained from a measurement should be reliable and valid (Noble, 2013).

In summary, self-assessment, if utilised properly, is a powerful clinical tool in terms of linking the results found in clinic to a person's experiences in the real world – about which psychoacoustic and physioacoustic measures can only make educated assumptions. That is to say, self-reporting answers significant clinical questions surrounding the day-to-day disability of hearing impairment (Noble, 2008).

1.5 Outcome Assessment

If the goal of outcome assessment is to verify the effectiveness of a particular treatment (in this case the treatment of hearing impairment), then it is imperative that the method of evaluation reflects the intent of that treatment (Boothroyd, 2007). Boothroyd (2007) suggests that there are four elements of aural rehabilitation; sensory management, instruction, perceptual training and counselling. As such, a number of different outcome assessments should be employed, in correspondence with the purpose of the aforementioned rehabilitative methods.

1.5.1 Sensory Management

Sensory management is intended to assist auditory function by aiding whatever natural structures are in place, such as the cochlea or auditory nerve. Therefore, any outcome assessment pertaining to sensory management should attempt to evaluate the success of the chosen management scheme. The most commonly prescribed method of sensory management within the practice of audiology is the use of hearing aids, and as such, specific measures have been developed in order to assess the output of a hearing aid in situ. The most common method of hearing aid verification is Real-Ear Measurement (REM), in which a small probe microphone is placed near the ear drum in order to evaluate Real-Ear Aided Gain (REAG), and compare it with the suggested output levels specified by the fitting prescription.

The REM is the most time efficient and dependable technique for validating hearing aid output performance (Valente & Valente, 2014), and has approximately a 3dB repeatability threshold (at a 95% confidence interval), meaning that changes made are accurate to within a 3dB variance. There is a significant amount of literature suggesting that accurately meeting the prescribed targets leads to more favourable outcomes for the user, both in terms of general sound quality and speech perception (Aazh, Moore, & Prasher, 2012;

Byrne & Byrne, 1986; Byrne & Cotton, 1988; Ching, Scollie, Dillon, & Seewald, 2010; Moore, Alcántara, & Marriage, 2001). Functional gain measures (such as unaided soundfield puretone threshold minus aided soundfield puretone threshold), by comparison, have a 15dB threshold for repeatability (at the same 95% confidence interval) (Hawkins, Montgomery, Prosek, & Walden, 1987; Valente & Valente, 2014). In saying that, however, Boothroyd's aforementioned sentiment that any outcome measure of a given treatment should be reflective of the intent of that treatment, is still applicable here. In which case functional assessments such as aided minus unaided spondee thresholds still have some validity, as the general goal of a hearing aid is to improve recognition of speech sounds. Subjective measures are similarly important. Examples of these are loud claps, rustling paper or rattling a spoon in a cup in order to subjectively assess loudness discomfort levels with common sounds that the client will likely be confronted with in daily life. Likewise, subjectively checking sound quality, sound balance, feedback and own-voice quality are essential components of outcome assessment.

Further to this, self-reporting is a final, but no less imperative aspect of outcome measure, as ultimately, in person centred care, the client's opinion of the treatment outcome holds far more weight than the opinion of the clinician (Cox, 2003). Vestergaard (2006) asserts that while objective assessments indicate the effectiveness of the amplification or technical features of a device, self-reporting is able to give a holistic representation of the overall treatment process. In the past this may have been conducted as an informal conversation between the client and clinician. However, this was not necessarily formally tracked or treated as scientific (Cox, 2003). By way of formalising this process, various self-report questionnaires have been developed based on normative data. Some examples of these are; the APHAB (Abbreviated Profile of Hearing Aid Benefit), which helps to evaluate remaining activity limitations a person may have; the SADL (Satisfaction with Amplification

in Daily Life), which looks at underlying aspects of how satisfied the client is as a result of their hearing aid, though without actually using the term “satisfaction”; and the HHIE (Hearing Handicap Inventory for the Elderly), which is designed to assess remaining participation restrictions for a client (Cox, 2003). Other examples of self-reported questionnaires are the DOSO (Device Oriented Subjective Outcome Scale), the MARS-HA (Measure of Audiologic Rehabilitation Self-Efficacy for Hearing Aids), and the IOI-HA (International Outcome Inventory for Hearing Aids).

1.5.2 Instruction

Boothroyd (2007) suggests there is evidence that formal instruction regarding HA use and management, as well as HA accessories, results in better function and greater use over time. That is to say, effective instruction leads to greater self-efficacy with HAs. Rebecca J. Kelly-Campbell and McMillan (2015) add that reasonable HA self-efficacy (HA-SE) is directly linked to HA satisfaction and overall HA outcomes. Furthermore, higher HA-SE leads to greater HA use (Hickson, Meyer, Lovelock, Lampert, & Khan, 2014; McMullan, Kelly-Campbell, & Wise, 2018). As the benefits of greater HA self-efficacy are possibly significant for the client, outcome measures have been developed to help the clinician evaluate their instruction, and so refine their methods until the desired result is achieved. These outcome measures are primarily self-report, due to the fact that instruction is focussed on self-efficacy more than perceptual or sensory management. An example of a questionnaire designed for this purpose is the MARS-HA, which assesses skill-based behaviours such as HA handling and use (e.g. changing batteries, insertion and removal etc), HA knowledge, and confidence in different listening environments (Smith & West, 2006).

1.5.3 Perceptual Training

Perceptual training (or auditory training) does not specifically target a particular function in and of itself, but instead it helps to facilitate better use of that function by aiding perceptual ability (Boothroyd, 2007). Sweetow and Palmer eloquently state that "hearing is only the first step in a cascade of events leading toward communication" (2005, p. 495). Taking both these notions into account insinuates that the true benefits of perceptual training may not become apparent instantaneously, but will become more perceptible as the user spends time developing the myriad of skills required for communication in real-life environments (Boothroyd, 2007). Regardless of the timeline of the intervention however, the literature makes it clear that there are measurable benefits to perceptual training, both in the short and long term (Ingvalson, Lee, Fiebig, & Wong, 2013; Zaballos, Plasencia, González, De Miguel, & Macías, 2016). Determining the success of the outcome of auditory perceptual training requires the perceptual skill of the client to be assessed in an appropriate way. This is to say, PTA would be an ineffective and invalid method of testing the outcome of speech in noise training, as it does not assess the perceptual skill of listening to speech in noise. Therefore, the outcome evaluation of perceptual training ideally requires a simulation of the listening situation that the intervention was designed to improve. Two such options in audiology are the speech in noise testing, and speech in quiet testing.

Speech in noise testing presents speech to the client with varying levels of background noise present. The goal of the test is to establish the signal to noise ratio at which the client achieves a 50% score (Wilson, McArdle, & Smith, 2007). Four commercially available examples of speech in noise tests are the Words-in-Noise Test (WIN), the Quick Speech-in-Noise Test (QuickSIN), the Bamford-Kowal-Bench Speech-in-Noise Test (BKB-SIN), and the Hearing in Noise Test (HINT).

The QuickSIN and WIN both use multitalker babble as a distracting noise, with the QuickSIN using Institute of Electrical and Electronics Engineers (IEEE) sentences as the test stimulus, while the WIN uses the Northwestern University Auditory Test No.6 (NU-6) word list instead. Interestingly, both the WIN and QuickSIN result in similar signal-to-noise performance ratios for clients with comparable hearing impairment (Wilson et al., 2007).

The BKB-SIN and HINT both utilise Americanised BKB sentences as test stimuli, which are short, with high levels of semantic and syntactic context, written to a first grade reading level, with multitalker babble as the distracter. Conversely, the HINT employs a speech-spectrum noise as background noise.

Speech in quiet testing is also a potentially important method of validating the result of auditory training, as it focusses on an individual's maximum ability to discriminate speech in isolation from other factors, such as background noise (McArdle & Hnath-Chisolm, 2014).

It is worth noting that while speech in quiet testing can function as a crosscheck for PTA, there is not necessarily any correlation between a listener's PTA and ability to understand speech in background noise (Kreisman et al., 2014). Therefore, in the context of the assessment of perceptual training, it must be emphasised that both speech in quiet and speech in noise testing serve purpose, and therefore should be used in conjunction with one another, not in opposition.

Finally, self-assessment can also be a useful tool for outcome evaluation of perceptual training, as it provides the clinician an insight into the client's personal, subjective experience with the intervention, rather than merely assuming the outcome based on objective measures (Boothroyd, 2007).

1.5.4 Counselling

Counselling is a significant role of the audiologist, as it focusses on an individual's quality of life, and their ability to participate in it, and helps participants to develop new and

complimentary methods of overcoming the communication breakdowns they experience (Hawkins, 2005). There is literature to show counselling to be a significant factor in HA success, Northern and Beyer (1999) report a 9% HA return rate for adults who do not attend counselling-based group aural rehabilitation, but only a 3% HA return rate for those who do attend (Hawkins, 2005).

As it is a more subjective method of intervention, however, the outcomes are widely variable and are largely determined by the characteristics of both the client and the clinician/counsellor, and their combined rapport with each another (Boothroyd, 2007). Citing Boothroyd's principle that assessment methods should reflect the intent of the intervention, subjective treatment therefore warrants subjective outcome measures, and consequently, self-assessment is potentially the only truly valid means of measuring the outcome of counselling.

1.6 Self-Assessment: PROMs

Self-assessment, or patient-reported outcome measures (PROMs) have the potential to be an effective and useful tool both to the researcher and to the clinician, provided that it is utilised in an appropriate context and manner (Noble, 2008). There are obvious benefits and downfalls to self-assessment, in that much of the validity of the results are based on highly subjective foundations, i.e. the client, and so can be manipulated by the client either positively or negatively (Noble, 2013). Several of these factors, both positive and negative, are discussed in further detail below.

1.6.1 Benefits

The preliminary advantage of a PROM is that it allows the measurement of an expression of personal experience and quality of life, which is difficult to evaluate objectively with any degree of certainty. These personal factors are complex and multifaceted, resulting from the combination of both biological, psychological, and

sociocultural influences (Dean, Orford, Staines, McGee, & Smith, 2017). As such, objective measures cannot possibly give insight into the intrinsic wellbeing of a person. However, self-reported, experiential information may not only allow perception of an individual's well-being (or lack thereof), but help develop a rehabilitation plan in response to that (Bentler, Kramer, & Sophia, 2000).

For any test to be valid, it must actually reflect the intent of the test (Bentler et al., 2000; Hyde, 2000). Therefore, by virtue of the nature of self-reporting, PROMs have high validity, provided that they are utilised in an effective context by the clinician (Bentler et al., 2000). This fact leads to the point that there is an onus on the clinician to be conscious of a client's health literacy and modulate the way in which outcome measures are administered based on the patient's needs and comprehension (Greywoode, Bluman, Spiegel, & Boon, 2009).

The reliability of an outcome measure is often reported in terms of a critical difference (CD) value. These are estimated based on the confidence interval of test re-test reliability data. CD scores help the clinician determine how reproducible the result of a particular questionnaire is, without said questionnaire having to sacrifice a balance of homogeneity and diversity within the test items themselves (Bentler et al., 2000; Hyde, 2000). Published PROMs should be based on normative data with established CD values, and also have a good level of reliability.

1.6.2 Drawbacks

PROMs allow the assessment of many things that cannot otherwise be assessed. Likewise, however, there are several limitations to what they may assess accurately, and still more barriers to their effectiveness even when they are a relevant method of assessment. For instance, questions are reliant on the assumption they are representative of the client's

context in life (Cox, 2003), as without this they cannot provide useful information to the clinician.

Greywoode et al. (2009) rightly note that a clinician should be cognizant of a patient's comprehension ability, and so adjust the way they present their material accordingly. While this stands true as an ideal, it highlights a second drawback of the use of self-reported measures; in order to ensure the validity of results of a questionnaire, the clinician should be heavily involved in administering it and scoring it, in order to mitigate the negative effects of poor health-literacy or comprehension, which in turn poses a significant administrative burden on the clinician.

Finally, and perhaps primarily, PROMs place a high level of burden on the client themselves. In order for a person to give valid responses they must be able to sufficiently understand the language of the PROM, and furthermore, must be able to apply the questions to their own context. A patient must therefore be able to accurately access memory of their own experiences, and if not, then they may provide answers that are missing, or that contain only partial information. Therefore, PROMs are reliant on a high level of health literacy and general cognition, and consequently may pose significant burden on the respondent, which in turn has the ability to significantly impact the validity and reliability of the results (Atcherson et al., 2011).

1.7 Health Literacy

The definition of health literacy is constantly evolving, as the role and standard of literacy in society changes, along with the advancement of the medical and scientific fields (Berkman, Davis, & McCormack, 2010). For the purpose of this thesis, however, several helpful definitions are drawn upon, as they highlight different aspects of health literacy. Ratzan (2001) cites the widely utilised bibliography compiled by Selden, Zorn, Ratzan, and Parker (2000, p. ix), which defines health literacy as "the degree to which individuals have

the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions". Ratzan also quotes Kickbusch (1997, p. 269), who suggested that "health literacy implies the achievement of a level of knowledge, person skills, and confidence to take action to improve personal and community health by changing personal lifestyles and living conditions". In addition to these, Nutbeam (2008, p. 2073), citing the work of Baker (2006), proposes a slightly simpler concept that health literacy is "the set of individual literacy capacities that act as a mediating factor in health and clinical decision-making". What all three definitions share thematically, is the fact that health literacy comes into effect through the personal context of an individual interacting with society. It therefore makes logical sense that health literacy on an individual level will be largely correlated to a person's societal context. As such, there are both direct and indirect associations between low literacy and a variety of undesirable health outcomes within a population (Nutbeam, 2008). Within the field of audiology, an individual's level of health literacy has the potential to affect the validity of self-reporting measures, and therefore the course of treatment for that individual.

1.7.1 At-risk Populations

Research has shown that several social groups are at a particular risk of having low health literacy. Parker (2000), shows that aging has a direct correlation with decreasing health literacy, and therefore seniors are among the most 'at risk' demographics (Murray, Hagey, Willms, Shillington, & Desjardins, 2008; Schecter & Lynch, 2011). This is of particular importance to the practice of audiology, as seniors account for one of the primary populations of those utilising audiological services. Other 'at risk' populations include immigrants, the unemployed, and those within a low-income demographic (Murray et al., 2008; Rudd, 2004;

Schecter & Lynch, 2011), as well as communities in rural areas (Quigley, Coady, Gregoire, Folinsbee, & Kraglund-Gauthier, 2009; Schecter & Lynch, 2011).

1.7.2 Improving Health Literacy

It is clear that there are, unfortunately, a myriad of barriers to health literacy across many demographics. Therefore, as the range of people affected by low health literacy is diverse, the different approaches used to improve health literacy must be equally as varied. Regular reading is positively linked to higher health literacy levels (Murray et al., 2008), and Schecter and Lynch (2011) note that more targeted reading activities lead to 52% higher scores for people aged 66 years or greater when assessed. On that basis, adult literacy programs could arguably be used to much greater effect than at present, however, the literature makes it clear that the issue of health literacy cannot be solved with a one-size-fits-all approach.

If adult literacy programs are indeed a suitable option for an individual, then the learners should be granted a significant level of control in terms of the structuring and content of the program, and have an overall participatory approach (Schecter & Lynch, 2011). Both Kagitcibasi, Goksen, and Gulgoz (2005) and Schecter and Lynch (2011) believe that content must also have significant connection to the everyday lives of the adults concerned, using literacy as a mediator in the cultural and social context of the learner (Purcell-Gates, Degener, Jacobson, & Soler, 2002). Schecter and Lynch (2011) note that much of the material designed to improve patient understanding and health literacy is, somewhat ironically, print-based media, which poses an issue for those who have a low literacy level. In this case, Davis et al. (1998) recommend that the use of videos can be an effective means of improving adult health knowledge, therefore mitigating some of the need for printed resources for those to whom print media is not an effective means of education. Indeed,

Greenberg (2001) found that even culturally appropriate brochures were ineffective for adults with low literacy levels, and therefore a better approach would be video. It is also valuable if an educator can acknowledge a learner's prior understanding, and use this as a base on which to build subsequent knowledge (Schecter & Lynch, 2011).

1.8 Readability

The working definition of readability used for the purpose of the study is the ease, speed and efficiency with which an individual may read, process and comprehend written language, based upon the characteristics of the written material (Pires, Cavaco, & Vigário, 2017).

1.8.1 Measurement

The concept and importance of readability has been present since the ancient Greeks, who realised that if a legal argument could not be understood by those listening, then it could hardly be persuasive (Collins-Thompson, 2013; Zakaluk & Samuels, 1988). In more recent times, however, the field of readability has expanded rapidly, with more than 200 readability formulas being produced since the 1970s (Crossley, Skalicky, Dascalu, McNamara, & Kyle, 2017). These formulas are designed to help predict the readability of text, based on syntactic complexity and lexical sophistication - which are normally calculated from the length of both words and sentences (Crossley et al., 2017). These measures can either be actual formula-based calculations (non-computational), or alternatively, many modern measurements are now deduced from computer-based algorithms in conjunction with machine learning (computational) (Collins-Thompson, 2013).

Though there are many different readability formulas in existence, some are more suited to the area of healthcare than others. Three such measurements are the Flesch–Kincaid

Grade Level Formula (F-K), Simple Measure of Gobbledygook (SMOG) and the FORCAST (Douglas & Kelly-Campbell, 2018).

1.8.2 Readability Measurements

The Flesch–Kincaid Grade Level Formula, (F-K) as it currently stands, was developed for the U.S. Navy in 1975, based on Flesch's 1948 Reading Ease Formula, and was intended to formulate a U.S. grade-level score (Douglas & Kelly-Campbell, 2018; DuBay, 2004; Kincaid, Fishburne, Rogers, & Chissom, 1975). The F-K calculation is based upon a 75% comprehension level, and calculates U.S. reading grades from 3-12 (D'Alessandro, Kingsley, & Johnson-West, 2001). The F-K has also become widely utilised due to its extensive availability. Fitzsimmons, Michael, Hulley, and Scott (2010), however, argue that the 75% comprehension level used in the reading level calculation leads to the F-K underestimating the true reading difficulty of a text (Douglas & Kelly-Campbell, 2018).

The SMOG formula is calculated based of a 100% comprehension level, compared with the 75% level of the F-K, and as such, has the potential to overestimate the true difficulty of a text. Fitzsimmons et al. (2010), therefore, recommend that the SMOG be favoured above other readability formulas when evaluating person-centred healthcare material, as it is more likely to lead to a more conservative estimation of an individual's health literacy. The SMOG also covers a far larger range of U.S. grade scores, with calculations allowing for a range of grade 3 through to grade 19 (Douglas & Kelly-Campbell, 2018). D'Alessandro et al. (2001) note that the accuracy of estimation for scores beneath the 6th grade reading level may be inhibited, however.

The FORCAST formula distinguishes itself from both the F-K and the SMOG in that it does not focus on a sentence-length measurement, but pays particular attention to functional literacy (Atcherson et al., 2011; DuBay, 2004). Formulas that are based on

sentence-length measurement have a distinct disadvantage when it comes to questionnaires, in that questionnaires often use incomplete sentences, or other nonnarrative, short texts (Atcherson et al., 2011; Douglas & Kelly-Campbell, 2018). As such, the FORCAST is particularly useful in the area of self-report measures.

1.8.3 Recommended RGL

There is ample literature to show that many individuals have very limited health literacy (Davis et al., 1998; Parker, 2000). This reality is often exacerbated by the fact that many health professionals frequently use medical jargon when communicating with their patients (Parker, 2000), and much health information is also written at a level far above the comprehension of an average client (D'Alessandro et al., 2001; Paasche-Orlow, Taylor, & Brancati, 2003; Svarstad, Bultman, Mount, & Tabak, 2003). As such, recommended reading grade levels for health materials have been established, in order that the maximum number of adults can read and understand health information. There is general agreement among researchers that this level should be between the 4th and 6th grade reading levels (Douglas & Kelly-Campbell, 2018; Matthews & Sewell, 2002; Wang, Miller, Schmitt, & Wen, 2013; Weiss, 2007), or at the very least it should not surpass the 6th grade RGL (Doak, Doak, & Root, 1996; Donald & Kelly-Campbell, 2016; Douglas & Kelly-Campbell, 2018; Friedman & Hoffman-Goetz, 2006; Yin, Forbis, & Dreyer, 2007). As previously mentioned in this chapter, the literature suggests that health literacy in the elderly population is especially low, due to decreasing cognitive capacity among other factors. Consequently, the recommended reading grade level is lowered for the aged demographic, recommending a reading grade level between the 3rd and 6th grades (Caposecco, Hickson, & Meyer, 2014).

The study of Douglas and Kelly-Campbell (2018, p. 66) assessed a wide range of PROMs used in audiology. Based on the FORCAST formula, they suggest that “a minimum

of an 8th US grade reading level was required in order to adequately read and comprehend the majority of the PROM sections, with the readability of many PROM sections exceeding the 10th to 11th grade". The means that there is a significant proportion of adults completing PROMs in audiology clinics that are not fully understanding the questions they are endeavouring to answer accurately.

1.8.4 Issues with Exceeding Recommended Level

Positive client-clinician relationship is centred around effective communication, mutual understanding and active participation from both parties. If a client's ability to effectively comprehend and engage with health materials, such as PROMs, is impeded, then none of the aforementioned aspects of positive relationship can be assumed. This is the case when PROMs exceed the recommended readability level health information, and therefore the desirable patient-centred care model inadvertently becomes unattainable (Douglas & Kelly-Campbell, 2018). The validity of any response is based upon the responder's comprehension of the asked question, and therefore a genuine, valid answer must be founded on genuine, valid comprehension. Both Atcherson et al. (2014) and DuBay (2004) further add that if the required literacy level of a document exceeds the literacy level of the reader, then there is a greater chance of the reader giving up on the task absolutely. In either instance, PROMs written at too high a level have a negative impact on person-centred care, and on the ability to achieve a desirable outcome for client and clinician alike.

1.9 Aim and Hypothesis

The over-arching aim of this study is to investigate whether revising PROMs so that they fall within the recommended RGL for health literacy will lead to more valid results than what the current PROMs produce. In order to assess this, there are two specific aims of this study:

1) To revise PROMs to meet RGL guidelines.
2) To assess whether there are differences in both results and reliability between the original and revised PROMs.

This follows on from the work of Douglas and Kelly-Campbell (2018), which showed that the overwhelming majority of all PROMs regularly used in the field of audiology specifically were above the recommended RGL for health literacy. These findings are consistent with other literature within the field of audiology (Atcherson et al., 2013; Atcherson et al., 2011; Rebecca J Kelly-Campbell et al., 2012), and also with literature within the general health sector (D'Alessandro et al., 2001; Hansberry et al., 2014; Kong & Hu, 2015).

Considering the current literature within the field, the following hypotheses for this study are:

1) Scores on the revised PROMs will be significantly different from scores on the original PROMs.
2) The revised PROMs will have a better test-retest reliability than the original PROMs.

Chapter 2 : Methods

2.1 Overview

Douglas and Kelly-Campbell (2018) showed that the overwhelming majority of PROMs exceed the recommended 6th RGL for health information when measured by several different readability measures, including the F-K, SMOG and FORECAST. The 6th RGL value is in reference to the grade level system used in U.S. education, meaning that health material should be comprehensible to a sixth-grade student. Based on this information, the aim of the present study was to investigate if revising a selection of PROMs, that were all above the recommended RGL for health information, would make a significant difference to the results the PROMs yielded. The PROMs selected for this study are listed later in this chapter.

2.2 Ethical approval

Before commencing this study, ethical approval was sought from the University of Canterbury. Following the initial application to the university ethics department, several small amendments were made, and the revised submission was subsequently approved for the study to go ahead.

2.3 Materials

Three PROMs were selected for revision based on the study by Douglas and Kelly-Campbell (2018). The PROMs chosen were: The Device-Oriented Subjective Outcome Scale (DOSO), the International Outcome Inventory for Hearing Aids (IOI-HA), and the Measure of Audiologic Rehabilitation Self- Efficacy for Hearing Aids (MARS-HA). These questionnaires were revised collaboratively by the researcher and supervisor of this thesis, and drafts were adjusted until the RGL was below 6 as determined by the in-program readability tool in Microsoft Word using the F-K readability formula, as recommended by the

literature. These amendments were discussed by the researcher and supervisor in order to ensure the content of the PROMs was not meaningfully changed from the original, validated material. Drafts were also piloted on a group of adults, to make sure they were understandable to members of the general public.

2.4 Participants

Sample size analysis, using G*Power software, indicated that 22 participants would be required to detect the desired effect size of $\eta^2 = .09$ (using an alpha level of .05 and a power level of .80). Five audiology clinics in the United States participated in this study. These clinics recruited consecutive hearing aid clients until the 22 participant requirement was achieved at minimum, though 45 participants were eventually included. The clinics were contacted via email, and the process was explained to them. The clinics invited eligible clients to participate in the study, and potential participants were referred to the researcher directly for all subsequent data collection and correspondence. Participant inclusion criteria were:

(1) Adults over the age of 18 years

(2) Fitted with at least one hearing aid, at least 1 year prior to data collection

(3) Able to read in the English language

These participants were chosen as a representation of the population of interest; adults with hearing impairment who have undergone intervention.

2.5 Procedures

The consenting participants were sent a study packet which included:

(1) A consent form, including a release of medical information and a postal address

(2) The first PROM (as determined by random assignment to order)

(3) A demographic form

(4) A postage-paid envelope to return the materials to the researcher

The participant demographic form included questions regarding age, gender, ethnic identity, HIHA history, education and occupation, along with a health literacy screening tool. Participant responses were recorded so that significant differences in results could be interpreted in the light of any demographic variables.

The participants were sent a mixture of two revised versions, and one original version of the PROMs. This method allowed test-retest reliability to be examined between the two revised versions, as well as any significant differences between the revised versions and the original to be reviewed and interpreted.

The participants were quasi-randomly assigned an order to complete the PROMs: The randomisation had the constraint that each of the orders had an equal number of participants assigned to them. The orders were:

(1) Original, revised, revised

(2) Revised, revised, original

(3) Revised, original, revised

2.6 Analyses

The results of the participant's responses were recorded in Microsoft Excel. These results were imported into the Statistical Package for Social Sciences (SPSS) for data analysis. The planned analyses were:

(1) Test-retest reliability: (a) Reliability analyses were performed in SPSS. (b) Critical difference scores were calculated following the discussion provided by Demorest and Walden (1984). The standard error of measurement (s_e) was calculated by multiplying the pooled standard deviation of the scores (s_x) by the square root of 1 minus the estimated reliability coefficient (r_{xx}): $s_e = s_x\sqrt{1 - r_{xx}}$. The critical difference (CD)

scores were calculated by multiplying the s_e by the square root of 2: $CD = \sqrt{2}s_e$.

(2) Mixed Model Univaraite analysis of covariance (ANCOVA): The scores on the PROMs were the dependent variable. The version of the PROM was the fixed factor (1 between and 1 within), demographic factors were the covariates. These were: age, score on the Short Assessment of Health Literacy (SAHL), better ear puretone average (BEPTA) (500, 1000, 2000, and 4000 Hz). Significant findings on the ANCOVA were followed up with repeated measures t-tests to determine which fixed factor accounted for significant differences.

Chapter 3 : Results

3.1 Participants

A total of 45 participants were recruited for this study (23 males and 22 females). All participants reported owning hearing aids (3 unilaterally fitted, 42 bilaterally fitted). Demographic information about the study participants is shown in Table 1.

Table 1. Demographic descriptive statistics for the study participants (N = 45).

Variable	Minimum	Maximum	Mean (SD)
Age (years)	54	72	63.62 (4.55)
Education (years)	4	16	11.29 (3.68)
Length of hearing impairment (years)	1	7	2.38 (1.68)
Length of hearing aid ownership (years)	1	4	1.55 (.76)
HHI Score	0	22	9.47 (5.50)
SAHL Score	6	18	12.35 (3.79)
BEPTA	6.00	70.00	37.80 (16.89)
WEPTA	20.00	78.00	43.86 (15.36)

HHI = Hearing Handicap Inventory, SAHL = Short Assessment of Health Literacy, BEPTA Better ear pure-tone average (of .5, 1, 2, and 4 kHz), WEPTA = worse ear pure-tone average (of .5, 1, 2, and 4 kHz).

Most participants exhibited a sensorineural hearing impairment that was sloping in configuration. Figures 1 and 32 show the mean audiograms for the right and left ears, respectively.

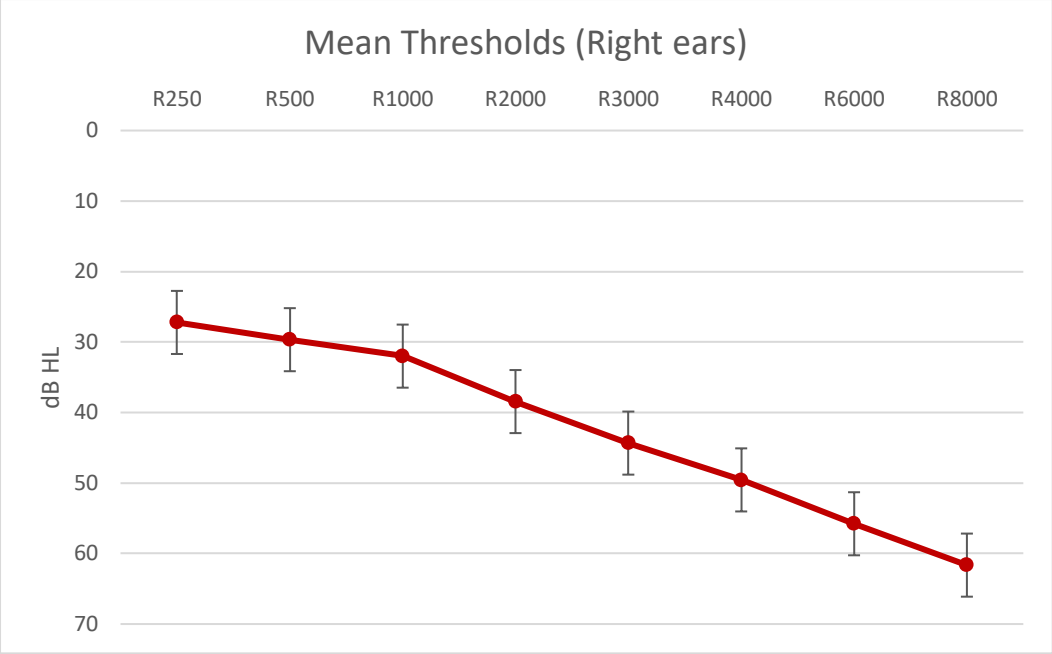

Figure 1. Means and standard errors for right ears for audiometric frequencies from 250 – 8000 Hz.

Figure 2. Means and standard errors for left ears for audiometric frequencies from 250 – 8000 Hz.

3.2 Hypothesis Testing

A series of univariate analyses of variance (ANOVA) were performed to test the study hypotheses. Prior to hypothesis testing, the data were assessed for normality. There were no significant outlying data points and the data were not significantly skewed or kurtotic. In addition, the data met the assumption of homogeneity of variance for each ANOVA. Partial eta squared (η^2) and Cohen's d effect sizes were calculated and reported.

3.2.1 Effect of order

A univariate ANOVA was performed to assess the possible effect of the order of the PROM administration on the study independent variables (age, HHI score, SAHL, BEPTA, WEPTA) and on the dependent variables (PROM scores). None of these ANOVA were significant ($p > .05$), as expected based on random assignment to order of administration.

3.2.2 Effect of health literacy

A univariate ANCOVA was performed to assess the possible effect of participant's SAHL score on the study dependent variables (PROM scores). There was a significant effect of SAHL on DOSO scores ($p < .001$), but not on the IOI scores ($p = .331$) or MARS scores (.514). Because there was a significant effect of SAHL on DOSO scores, SAHL was used as a covariate in those analyses.

3.2.3 Effect of PROM version

Each participant completed the PROMs three times in a randomized order; one administration of the original PROM and two administrations of the revised PROM (Revised 1 and Revised 2). It was hypothesised that there would be a significant difference between

the original version and the Revised 1 version of the PROMs and that there would not be a significant difference between the two administrations of the revised versions of the PROMs.

3.2.3.1 IOI

A repeated measures multivariate analysis of variance (RM-MANOVA) was performed to assess the possible effect of IOI version on participant scores. For each analysis, the assumptions of normality and homogeneity of variance/sphericity were met. There was a significant multivariate main effect of version on the IOI scores: $F(2,42) = .045$. The RM-MANOVA was followed up with repeated measures ANOVA for each IOI item and pairwise comparisons for each administration of the IOI. The means and standard errors for each administration of the IOI items and total score are shown in figure 3.3.

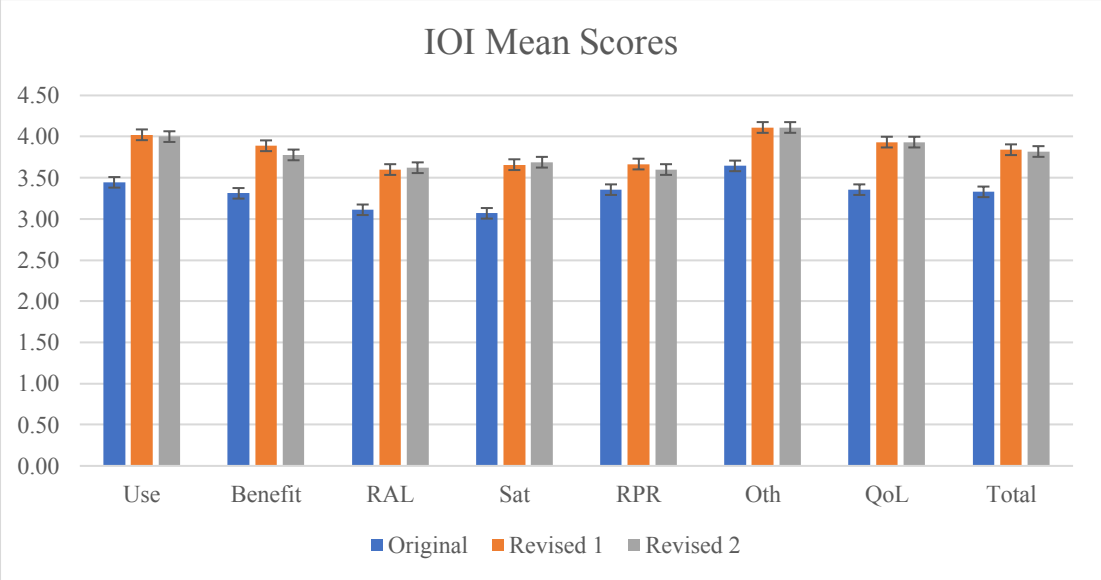

Figure 3. Means and standard errors of the three administrations of the International Outcome Inventory (IOI). RAL = Residual Activity Limitations, Sat = Satisfaction, RPR = Residual Participation Restrictions, Oth = Impact on Others, QoL = Quality of Life.

The RM-ANOVA revealed a significant main effect of version on IOI-Use scores: $F(2,86) = 6.9$, $p = .01$, $\eta^2 = .14$. Pairwise comparisons indicated there was a significant

difference between the original and revised versions ($p < .001$, $d = .87$); however, there was no significant difference between the two revised administrations ($p = .66$, $d = .04$).

The RM-ANOVA revealed a significant main effect of version on IOI-Benefit scores: $F(2,86) = 3.1$ $p = .04$, $\eta^2 = .07$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .93$); in addition, there was a significant difference between the two revised administrations ($p = .02$, $d = .21$).

The RM-ANOVA revealed a significant main effect of version on IOI-Residual Activity Limitation scores: $F(2,86) = 7.09$, $p = .001$, $\eta^2 = .14$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = 1.17$); however, there was no significant difference between the two revised administrations ($p = .56$, $d = .04$).

The RM-ANOVA revealed a significant main effect of version on IOI-Satisfaction scores: $F(2,86) = 7.80$, $p = .001$, $\eta^2 = .15$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .1.3$); however, there was no significant difference between the two revised administrations ($p = .32$, $d = .04$).

The RM-ANOVA revealed there was no significant main effect of version on IOI-Residual Participation Restriction scores: $F(2,86) = .64$, $p = .53$, $\eta^2 = .01$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .51$); however, there was no significant difference between the two revised administrations ($p = .08$, $d = .12$).

The RM-ANOVA revealed a significant main effect of version on IOI-Impact on Others scores: $F(2,86) = 8.75$, $p < .001$, $\eta^2 = .17$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .65$); however, there was no significant difference between the two revised administrations ($p > .99$, $d = 0$).

The RM-ANOVA revealed there was no significant main effect of version on IOI-Quality of Life scores: $F(2,86) = 1.12$, $p = .33$, $\eta^2 = .03$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .74$); however, there was no significant difference between the two revised administrations ($p > .99$, $d = 0$).

The RM-ANOVA revealed there a significant main effect of version on IOI-Total scores: $F(2,86) = 23.43$, $p < .001$, $\eta^2 = .35$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = 2.1$); however, there was no significant difference between the two revised administrations ($p = .17$, $d = .10$).

3.2.3.2 DOSO

A repeated measures MANOVA was performed to assess the possible effect of DOSO version on participant scores. For each analysis, the assumptions of normality and homogeneity of variance/sphericity were met. As stated previously, SAHL was used as a co-variate in these analyses. There was a significant multivariate main effect of version on the DOSO scores: $F(2,42) = .03$. The RM-MANOVA was followed up with repeated measures ANOVA for each DOSO scale and pairwise comparisons for each administration of the DOSO. Means and standard errors were calculated for the original and revised administrations of the DOSO. These results are shown in figure 3.4.

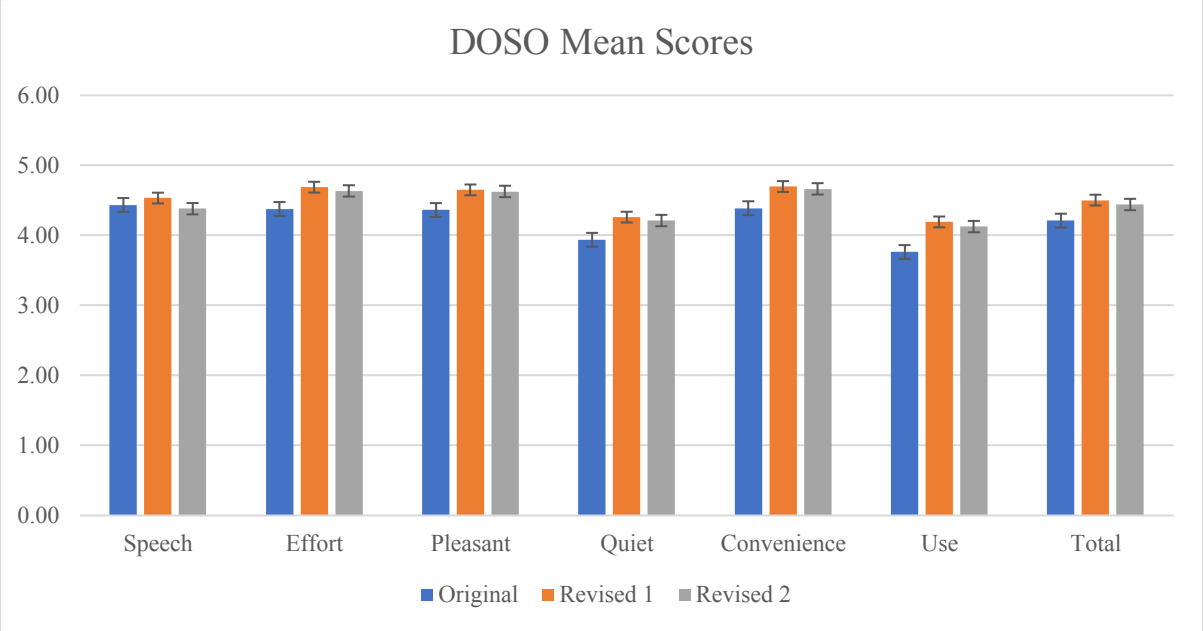

Figure 4. Means and standard errors of the three administrations of the Device-Oriented Subjective Outcome (DOSO) questionnaire.

The RM-ANOVA revealed a significant main effect of version on DOSO-Speech scores: $F(2,86) = 4.1$, $p = .02$, $\eta^2 = .09$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .36$); however, there was no significant difference between the two revised administrations ($p = .21$, $d = .19$).

The RM-ANOVA revealed a significant main effect of version on DOSO-Effort scores: $F(2,86) = 13.6$, $p < .001$, $\eta^2 = .24$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .60$); however, there was no significant difference between the two revised administrations ($p = .06$, $d = .12$).

The RM-ANOVA revealed there was no significant main effect of version on DOSO-Pleasantness scores: $F(2,86) = 1.89$, $p = .15$, $\eta^2 = .24$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .72$); in addition, there was a significant difference between the two revised administrations ($p = .04$, $d = .05$).

The RM-ANOVA revealed there was no significant main effect of version on DOSO-Quiet scores: $F(2,86) = 7.22$, $p = .001$, $\eta^2 = .14$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .54$); in addition, there was a significant difference between the two revised administrations ($p = .004$, $d = .07$).

The RM-ANOVA revealed there was no significant main effect of version on DOSO-Convenience scores: $F(2,86) = 14.36$, $p < .001$, $\eta^2 = .25$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .66$); in addition, there was a significant difference between the two revised administrations ($p = .03$, $d = .04$).

The RM-ANOVA revealed there was no significant main effect of version on DOSO-Use scores: $F(2,86) = 10.04$, $p < .001$, $\eta^2 = .19$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = .98$); in addition, there was a significant difference between the two revised administrations ($p = .04$, $d = .13$).

The RM-ANOVA revealed there was no significant main effect of version on DOSO-Total scores: $F(2,86) = 20.70$, $p < .001$, $\eta^2 = .32$. Pairwise comparisons indicated there was a significant difference between the original and revised versions ($p < .001$, $d = 1.15$); in addition, there was a significant difference between the two revised administrations ($p < .001$, $d = .36$).

3.2.3.3 MARS-HA

A repeated measures MANOVA was performed to assess the possible effect of MARS-HA version on participant scores. For each analysis, the assumptions of normality and homogeneity of variance/sphericity were met. There was a significant multivariate main

effect of version on the MARS-HA scores: F (2,42) = .002. The RM-MANOVA was followed up with repeated measures ANOVA for each MARS-HA scale and pairwise comparisons for each administration of the MARS-HA. Means and standard errors were calculated for the original and revised administrations of the MARS-HA. These results are shown in figure 3.5.

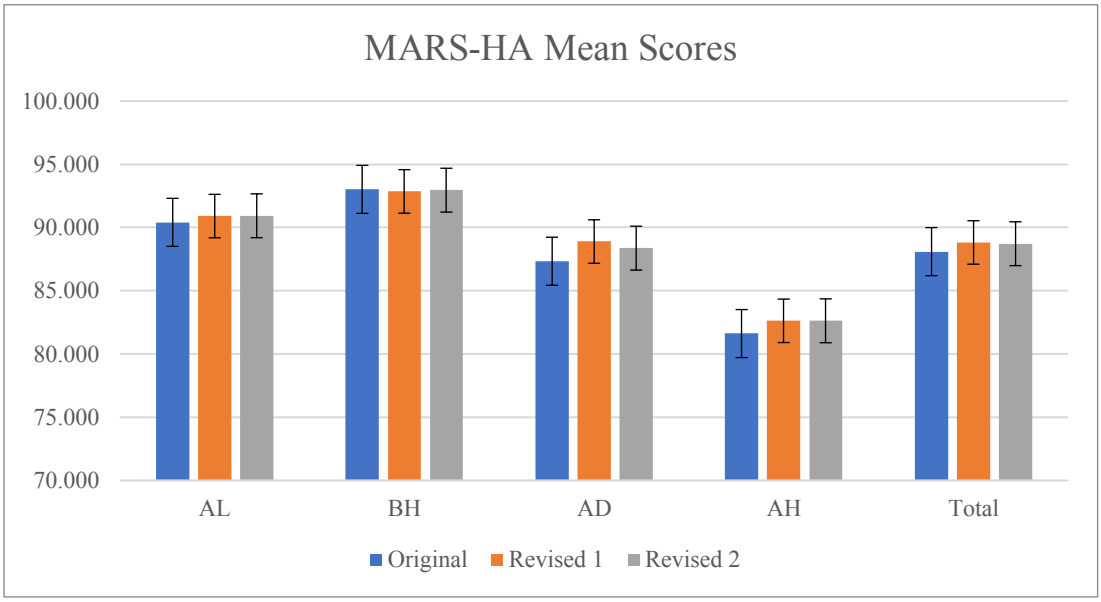

Figure 5. Means and standard errors of the three administrations of the Measure of Audiologic Rehabilitation Self- Efficacy for Hearing Aids (MARS-HA) questionnaire. AL = Aided Listening, BH = Basic Handling, AD = Adjustment, AH = Advanced Handling.

The RM-ANOVA revealed a significant main effect of version on MARS-HA-Aided Listening scores: F (2,86) = 28553.06, p = .02, η^2 = .99. Pairwise comparisons indicated there was a significant difference between the original and revised versions (p < .001, d = .35); however, there was no significant difference between the two revised administrations (p = .63, d = .03).

The RM-ANOVA did not reveal a significant main effect of version on MARS-HA-Basic Handling scores: F (2,86) = 1.45, p = .24, η^2 = .03. Similarly, pairwise comparisons did

not reveal a significant difference between the original and revised versions (p =.73, d = .02); or between the two revised administrations (p = .63, d = .01).

The RM-ANOVA revealed a significant main effect of version on MARS-HA-Adjustment scores: F (2,86) = 6.12, p = .003, η^2 = .13. Pairwise comparisons indicated there was a significant difference between the original and revised versions (p < .001, d = 1.6); in addition, there was a significant difference between the two revised administrations (p = .007, d = .13).

The RM-ANOVA revealed a significant main effect of version on MARS-HA-Advanced Handling scores: F (2,86) = 6.39, p = .003, η^2 = .13. Pairwise comparisons indicated there was a significant difference between the original and revised versions (p < .001, d = .19); however, there was no significant difference between the two revised administrations (p = .21, d = .05).

The RM-ANOVA revealed a significant main effect of version on MARS-HA-Total scores: F (2,86) = 3.19, p = .04, η^2 = .07. Pairwise comparisons indicated there was a significant difference between the original and revised versions (p =. 002, d = .18); similarly, there a significant difference between the two revised administrations (p = .021, d = .02).

3.2.4 Summary of results and effect sizes

It was hypothesised that there would be significant differences in IOI item scores between the original and revised version, but there would not be significant differences between the scores on the two revised administrations. The results of the ANOVA supported this hypothesis for all IOI items except Benefit. Cohen's d effect sizes associated with the significant differences ranged from .51 to 2.1. Cohen's d effect sizes associated with the non-significant differences ranged from 0 to .21.

The results of the DOSO ANOVA followed the expected pattern for only the Speech and Effort scales. For all other scales, there was a significant difference between the original and revised versions as well as between the two administrations of the revised DOSO. Cohen's d effect sizes associated with the significant differences ranged from .04 to .98. Cohen's d effect sizes associated with the non-significant differences ranged from .06 to .36.

The results of the MARS-HA ANOVA revealed that the expected pattern was found for the Aided Listening and Advanced Handling scales. There were no significant differences between the original and revised versions or between the two administrations of the revised version for the Basic Handling scale. There were significant differences between the original and revised versions as well as the two administrations of the revised version for the Adjustment scale. Cohen's d effect sizes associated with the significant differences ranged from .02 to .1.6. Cohen's d effect sizes associated with the non-significant differences ranged from .01 to .05.

Chapter 4 : Discussion

4.1 Introduction

The intent of this study was to investigate if revising PROMs in order to lower their RGL would illicit different results to the original versions, and also if the resulting revised PROMs would have better test-retest reliability than the originals. The PROMs chosen for revision were the DOSO, the IOI-HA, and the MARS-HA. Participants were recruited from five different clinics in the United States, and completed one copy of the original PROM, along with two copies of the revised version over a six-week period. The results indicate that the study was statistically over-powered. Though the trends in the data are encouraging, with regard to the hypotheses, there is too much variation in the results to conclude that the results support the hypotheses categorically. This chapter will discuss the results and their clinical implications in relation to the literature.

4.2.1 Effect of order

As shown in the previous chapter, the data showed no effect of order on the results. This result was as expected, due to the randomised order the PROMs were distributed in. The effect of order was also potentially limited via the participation selection process. Because all participants were HA users already, at least some level of familiarity with PROMs can be assumed, as all had some experience in audiology clinics. There should also have been no significant learning effect in the course of this study, as any potential learning effect would have happened during their previous experience with PROMs. Though the likelihood of either of these things happening in reality was doubtful, it is worth noting as a possible limiting factor.

4.2.2 Effect of health literacy

A univariate analysis of covariance (ANCOVA) revealed no significant effect of health literacy on either the IOI-HA or MARS-HA. However, a significant effect of health literacy was present with the DOSO. This effect was both interesting and unexpected, in that it clearly applied to the DOSO ($p < .001$), but equally as clearly did not a apply to either the IOI or MARS scores ($p = .331$), and ($p = .514$) respectively. The results were reanalysed in order to rule out any human error in their calculation, though this was not found to be the case.

It is possible that this result was influenced, at least partially, by the health literacy screening tool used in the demographic data collection phase, the SAHL. The SAHL is designed as a simple literary association tool, functioning as a screening indicator of a person's health literacy. However, in that the task only involves correctly associating one health-related word with another (between two alternatives), this means that with the exception of the "I don't know" response option, there is a 50 percent chance of the correct answer being guessed, regardless of health literacy. Furthermore, and perhaps more importantly, there is little account taken of the individual's general literacy skills, as opposed to health-specific literacy. Therefore, an individual's ability to answer questions related to health literacy may be limited in their ability to do so if they have poor general literacy. Wolf, Feinglass, Thompson, and Baker (2010) infer that with many literacy assessments, if a person's literacy is low enough, it becomes very difficult to interpret the relationship between literacy and the variable of interest, as literacy and the other variable may become independent past a certain threshold. Therefore, in some cases, it could be that the SAHL was not specific or detailed enough to accurately relate responses with genuine differences in health literacy that did, in fact, influence the results in some way. It is possible that this comes into effect when completing sentence-based PROMs, such as the DOSO.

This does not account for why this response occurred with the DOSO, but not with the other PROMs, and the true reason is unknown. This issue will be discussed further later in this chapter under “study limitations”.

4.2.3 Effect size

Upon analyzing the raw data, it quickly became apparent that the study was overpowered, causing the statistical significance to support the null hypothesis concerning test-retest reliability between the two revised versions of each PROM. However, upon further inspection, it was clear that the statistical significance was not an accurate expression of the trends within the data. The effect sizes for each section were, in fact, more representative of the underlying results.

The overpowering occurred as a result of having more than twice the number of participants necessary to find statistically significant results, causing even minimal variations in the data to appear significant. Even with the application of a Bonferroni correction, many instances in the results still indicate a significant difference between the two revised versions. However, if the revised/revised Cohen’s d values are observed and compared with the Cohen’s d values of the original/revised versions, there is a stark contrast in the magnitude of the numbers. This will be expanded on later in this chapter, when the results from the PROMs are discussed individually.

4.3 Effect of PROM version

As previously specified, each participant completed the PROMs three times in a randomized order, with one administration of the original PROM and two administrations of the revised PROM. It was hypothesised that there would be a significant difference between the original and the first administration of the revised version, and that there would be no

significant difference between the two instances of the revised versions of the PROMs, indicating greater test-retest reliability.

4.3.1 IOI-HA

The IOI-HA, overall, was very supportive of both hypotheses. As shown in the results chapter, all sections, barring one, indicated that the scores on the revised PROMs were significantly different to the scores on the original PROMs, and that the revised PROMs also had better test-retest reliability than the original PROMs. The only exception to these results was the section on hearing aid benefit. This section showed a statistically significant difference between the original and revised versions, as expected, but additionally revealed a statistically significant difference between the two revised versions. This, therefore, does not seem to support the second hypothesis, that there would be no significant difference between the two revised versions, representing improved test-retest reliability compared to the original version.

However, comparing the Cohen's d values reveals that while there was a difference of .93 of a standard deviation between the original and revised scores for this section, there was only a difference of .21 of a standard deviation between the two revised versions. Rice and Harris (2005) cite the work of Cohen (1969, 1988), advising that differences of .2 of a standard deviation between two variables are virtually imperceptible, while differences of .8 are clearly noticeable, and indicate as large a difference between two means as is common. Contrasting the trends in the data in this way indicates that the test-retest reliability of the revised version was, in fact, much improved compared to the original, even if the statistical significance indicates otherwise.

4.3.1.1 Limitations (IOI-HA specific)

There were several limitations to this study that influenced the results in this way, and these will be discussed in greater detail later in this chapter. Other than the general limitations of this study, this particular result could potentially also be due to the nature of the question, as it frames a past-tense issue; "Think about the one situation where you most wanted to hear better, before you got your hearing aid(s)", with a present-time question; "Over the last two weeks, how much has the hearing aid helped in that situation?". DuBay (2004) recommends that good readability is associated with the use of simple words, active voice, and the present tense. On this basis, even though the questions are simple and utilise an active voice, and though the RGL itself is within the recommended limits for health literacy, the changing tense could potentially be confusing to those with lower general literacy skills. Overall, however, the "total scores" section showed that the revised IOI-HA supported both hypotheses, revealing a significantly different overall score between the two versions (d = 2.1), and no significant difference between the two original versions (d = .10), indicating improved test-retest reliability.

4.3.2 DOSO

The DOSO was less outwardly supportive of the hypotheses than the IOI-HA was, as the same statistical overpowering phenomenon that was present in the benefit section of the IOI-HA was even more pronounced in the DOSO. The speech and effort sections supported both hypotheses, with significant differences between the original and revised versions, without there being any significant difference between the two revised versions. The subsequent five sections; pleasantness, quiet, convenience, use and total score, supported the first hypothesis, but rejected the second by additionally revealing a significant difference between the two revised versions.

Yet again, however, observing the effect size between the two categories (original/revised and revised/revised), give further insight into the trends in this data. The section measuring pleasantness revealed a difference between the means original and revised versions of d = .72 standard deviations, while conversely, the difference between the two revised versions was d = .05 standard deviations. This pattern was reiterated throughout the remaining categories, with *quiet* contrasting a difference of d = .54 with d = .07 between original/revised and revised/revised respectively. *Convenience* contrasted d = .66 with d = .04 between original/revised and revised/revised respectively. The *use* category contrasted d = .98 with d = .13 between original/revised and revised/revised respectively. Finally, *total scores* contrasted d = 1.15 with d = .36 between original/revised and revised/revised respectively. The difference between the two revised versions is consistently diminutive in comparison to the difference between the original and revised versions. With this in mind, the trends in the data are encouraging in regard to the difference made by the revision to the PROMs, even if they cannot be defined as statistically significant.

4.3.2.1 Limitations (DOSO specific)

As with the IOI-HA, the pattern of results indicates that the inflated number of participants in this study has, indeed, caused the study to be overpowered. This means that though the majority of the results seem to support the null of the second hypothesis, regarding the comparison of the two revised versions, this may not be an accurate representation of the data. Whether this is the case or not, however, remains speculative until further investigation can be completed.

4.3.3 MARS-HA

The MARS-HA yielded the least consistent results of the three PROMs, revealing some sections that followed the expected pattern of results, some that showed significant differences between both the revised versions as well as between the original and revised, and some that showed no difference between either version.

The *aided listening* scores followed the expected pattern of results set out in the hypotheses, as did the *advanced handling* section. Both the *adjustment* and *total* scores revealed a significant difference between the original and revised versions, as well as between the two revised versions. As with the previously discussed PROMs, these results follow a similar pattern of having far larger Cohen's d values separating the means of the original and revised versions, than those separating the two revised versions, indicating a greater difference between the two versions than the statistical significance gives credit for. Conversely, due to the nature of adjustment, it is possible that an individual's perception of this aspect with vary from day to day, and so, could foreseeably change during the two-week intermission between two administrations of the PROMs.

The most surprising result was that there was virtually no difference between either version for the *basic handling* section of the MARS-HA, with Cohen's d values of .02 and .01 for the original/revised and revised/revised comparisons respectively. This particular result is possibly due to the fact that the participants were all experienced HA users, and therefore had likely had many of these more basic terms reiterated to them many times over the course of their HA experiences. If they were already confident with the terms associated with basic handling, then revising the wording will make little to no difference to the participant. In the same way, the responses of someone with a high level of health literacy would be less affected by the version, as neither the original nor the revised versions would obscure the meaning of the question to them.

4.3.3.1 Limitations (MARS-HA specific)

As mentioned, in the case of the MARS-HA, the nature of the questions themselves may have been as influential in regard to the inconsistency of the results as the fact that the study was overpowered. In the case of the *basic handling* section, the fact that participants would, plausibly, be familiar with the language of this particular section specifically limits its ability to assess the effect of the revision effectively. In the same way, if a person is very familiar with how to change the oil in a car, then they may not benefit from someone explaining the process to them in plain language. Though at the same time, specific knowledge of a single system is not necessarily representative of a good overall understanding of how a car works. Conversely, in the absence of the assumption participant familiarity, it would also be possible for an individual who did not fully comprehend the original version to achieve the same score by guess-work alone. Though this scenario is unlikely to be represented in such a large group, the nature of breaching the recommended RGL is that participant guess-work must be expected in the absence of comprehension, and therefore this reality much be entertained as a possibility.

In the case of the *adjustment* section, though there was a marked difference between the original and revised scores (d = 1.6) and, comparably, a minute difference between the two revised versions (d = .13), the variance was still statistically significant. As established, this is almost certainly due to the power of the study as the result of having over double the required number of participants. Regardless of this fact, however, it is possible that this particular section was also influenced by the nature of the section. Many audiologists will have experienced a client being unhappy with the sound of their hearing aids, adjusting the aids to suit the client's requests, only to find that the new settings are merely a reversion back to the same settings that the client was unhappy with two weeks prior. This is to say that,

often, an individual's perception of how well they are adjusting to the sound of hearing aids can fluctuate on a daily basis, depending on their current listening requirements. On this basis, the variance in the answers provided by participants between the two revised versions could well be representative of a change in preference due to natural fluctuation, not merely that the revised version still produces inconsistent results.

4.4 Clinical Implications

"The effectiveness of clinical intervention is ultimately contingent upon the client's appraisal of his or her problems, the subjective nature of self-report is not only appropriate, it is focal" (Erdman, 1993, p. 308). The use of PROMs in a clinical setting is beneficial to both clinician and client, as they not only provide information that is unattainable by objective methods, but also promote client participation (Erdman, 1993). The obvious implication that this study intends to highlight, is that if a client's ability to correctly comprehend PROMs is jeopardized, then so too is the validity of their results, so too the possibility for the client to engage in their own care, and therefore, so too the development of the patient centered care model (Atcherson et al., 2011; Douglas & Kelly-Campbell, 2018; Gilligan & Weinstein, 2014).

The results of this study depict an overarching trend that revising the wording of the current published PROMs in use within the field of audiology to within the recommended RGL for health-related materials is effective in producing more consistent, valid, and clinically useful results. The implication of this is, first and foremost, clear that further research is warranted in order to more definitively substantiate this these results. Secondly, that in accordance with the current shift towards patient centered care, the publishers and developers of the PROMs currently in use must consider the revision of their materials. These results also imply that the current results being achieved with PROMs in clinical settings will often have low validity if completed without some level of supervision by the clinician. For

the clinician it is worth noting here, additionally, that readability is but a single element of PROM comprehension, and it is imperative that the clinician is aware of the comprehension levels of their clients, and tailor the way PROMs are administered accordingly (Greywoode et al., 2009). Much as in other arenas in life, communication is imperative.

4.5 Limitations

There were several ways in which this study was limited, and several of these have already been discussed. As mentioned previously, the most obvious of these is that the study was overpowered. Though the instance of having twice the number of participants necessary to make the study statistically significant as determined by the G*Power software meant that the results had high validity, it also meant that the true significance of the results were less obvious than if there had been fewer participants. The observation of the Cohen's d values were indicative of this, as they underscored the comparative differences between the original and revised versions as well as between the two iterations of the revised version. Despite the fact that Cohen's d tells a different story, there is no way to draw a definitive statistically valid conclusion based on these results. Inferences can be made, connections can be drawn, but other than for the IOI-HA, no statistical conclusion can be drawn without further research.

Additionally, no information on first language use was recorded. Though being able to read in English was a prerequisite for participation in the study, there was no record of native language or the amount that the participant spoke in English if it was not their native language. This could potentially have had a bearing on the individual's general literacy, as separate from health literacy alone.

Continuing in this vein, though a health literacy screening tool was used (the SAHL), there was no general literacy assessment. In that health literacy is encompassed by an individual's general literacy skills, and is therefore either limited or enhanced by it, it would

have made sense to assess this alongside the SAHL, as this could potentially have a marked effect on a participant's results.

The test-retest reliability of the revised version was assessed by including two copies over the course of the study. However, it is hard to make accurate comparisons between the test-retest reliability of the two versions when there was no formal assessment of the original version itself. A future study would benefit from including two presentations of the original version, so that an individual test-retest measurement could be made, and subsequently compared with that of the revised version.

The revised versions of the PROMs were based off Microsoft Word's built-in readability calculator, which calculates using the F-K readability formula, and reports is as F-K RGL. As stated, the revised versions were re-worded until they fell within the recommended RGL for health-related materials according to Microsoft Word's readability calculator. The limitation here is that many readability tests rely on a minimum of one hundred words per section in order to calculate accurate estimates of the true RGL (Douglas & Kelly-Campbell, 2018). In the case of the PROMs used in audiology, having one hundred-word questions would not be feasible, and therefore the accuracy of the readability calculator may be impaired.

Finally, it is possible that the timing of the administrations had an effect on the results. In that the participants had to complete each of the IOI-HA, the DOSO and the MARS-HA in each instance, a potential limitation could be that the participant's apathy towards the PROMs could gradually increase from the first to third PROM within each administration, and more still from the first administration to the final one six weeks later after having already completed a set of three PROMs twice already. This would mean that the results became less representative of the individual's true responses when at minimum level of indifference.

Similarly, given that each PROM focuses on a slightly different aspect of aural rehabilitation, there could be optimum presentation intervals based upon the intent of the individual PROM. For instance, it could be optimum to present the DOSO at two-week intervals in order for the client to have an adequate amount of time to adjust to changes pertaining to their device, while the IOI-HA may be optimized for six-month intervals. It is very difficult to assess the effect of this limitation, or design a study to assess the optimum presentation intervals without, paradoxically, using some form of self-reported outcome measure.

4.6 Future Research

If nothing else, this study outlines the need for future research, both into the development of new PROMs, and particularly, the revision of current ones. The results from this study indicate there is measurable value in lowering the RGL of PROMs to within the recommended levels for health literacy, and there is an opportunity for future studies to mitigate the limitations outlined above and make a clear case for the revision of all PROMs currently in use in audiology. Future research should take into account that it is not only word length or syntactic complexity that influence readability. Cosmetic factors such as font, text size, the amount of white space on the page, general attractiveness, along with the inclusion or exclusion of diagrams and pictures contribute significantly to the overall readability of a PROM (Douglas & Kelly-Campbell, 2018; Rebecca J Kelly-Campbell et al., 2012; Zraick, Atcherson, & Brown, 2012). Future researchers and/or clinicians must also be aware that many people will commonly experience embarrassment in regard to having low literacy skills, and therefore may try to disguise the fact that they are having issues comprehending PROMs (Douglas & Kelly-Campbell, 2018; Friedman & Hoffman-Goetz, 2006). Moreover, El-Daly, Ibraheim, Rajakulendran, Culpan, and Bates (2016) suggest that even individuals that *do* have a high level of health literacy will avoid clarifying queries they may have, so

that they will not appear uneducated to the clinician (Douglas & Kelly-Campbell, 2018). In both instances, care and tact must be employed in order to mitigate these situations and ensure accurate understanding is confirmed wherever possible.

4.7 Conclusion

This study aimed to investigate the effects of revising PROMs to within the recommended RGL for health-related materials. Three published PROMs were revised to within the recommend levels, according the F-K readability formula, and were trialed on experienced HA users in audiology clinics, in comparison with the original versions of PROMs. The scores of the revised versions were analyzed and compared to those of the original versions, and also compared to a second copy of the revised version in order to assess test-retest reliability. It was hypothesized that the scores derived from the revised PROMs would be significantly different to the scores derived from the original versions, and also that the scores from the repeated revised version would not be significantly difference to the initial revised administration of the PROMs. The results, statistically, were largely inconclusive, with some sections appearing to support both hypotheses, some only supporting one hypothesis of the two, and one section that supported neither, based on statistical significance. This which was most likely due to the large number of participants, and subsequent overpowering of the study. However, the trends in the data were very supportive of both hypotheses, with the majority of the sections revealing a large difference between the means of the score of the original and revised versions, and a very small difference between the two administrations of the revised version.

Though the results are statistically murky, and underscore the need for future research, the trends revealed by this study are encouraging in regard to the hypotheses. The results are a preliminary indication that if PROMs are revised to within the recommended

levels for health-related materials, then clients will give more clinically valid, consistent answers than what they do with the current versions.

Boothroyd (2007) insists that in the interest of achieving an optimum result for patients, sensory management cannot be relied on by itself. Similarly, the implications of a patient misunderstanding health-related information are serious (Wang et al., 2013), and therefore significant consideration should be given to the revision of PROMs, even on this basis alone. In the interest of continuing the trend of becoming more patient orientated (Cox, 2005; Gilligan & Weinstein, 2014), and taking into account the implications to the client's wellbeing, there is a strong case for the revision of PROMs in the field of audiology.

References

Aazh, H., Moore, B. C. J., & Prasher, D. (2012). The accuracy of matching target insertion gains with open-fit hearing aids. *American Journal of Audiology, 21*(2), 175-180. doi:10.1044/1059-0889(2012/11-0008)

Atcherson, S. R., DeLaune, A. E., Hadden, K., Zraick, R. I., Kelly-Campbell, R. J., & Minaya, C. P. (2014). A computer-based readability analysis of consumer materials on the american speech-language-hearing association website. *Contemporary Issues in Communication Science and Disorders, 41*(Apr), 12-23.

Atcherson, S. R., Richburg, C. M., Zraick, R. I., & George, C. M. (2013). Readability of questionnaires assessing listening difficulties associated with (central) auditory processing disorders. *Language, Speech, and Hearing Services in Schools, 44*(1), 48. doi:10.1044/0161-1461(2012/11-0055)

Atcherson, S. R., Zraick, R. I., & Brasseux, R. E. (2011). Readability of patient-reported outcome questionnaires for use with persons with tinnitus. *Ear and Hearing, 32*(5), 671-673. doi:10.1097/AUD.0b013e3182134654

Baker, D. W. (2006). The meaning and the measure of health literacy. *Journal of General Internal Medicine, 21*(8), 878-883.

Bentler, R. A., Kramer, & Sophia, E. (2000). Guidelines for choosing a self-report outcome measure. *Ear and Hearing, 21*(4 Suppl), 37S-49S.

Berkman, N. D., Davis, T. C., & McCormack, L. (2010). Health literacy: What is it? *Journal of Health Communication, 15*(S2), 9-19.

Boothroyd, A. (2007). Adult aural rehabilitation: What is it and does it work? *Trends in Amplification, 11*(2), 63-71. doi:10.1177/1084713807301073

Byrne, D., & Byrne, D. (1986). Effects of frequency response characteristics on speech discrimination and perceived intelligibility and pleasantness of speech for hearing-impaired listeners. *Journal of the Acoustical Society of America, 80*(2), 494-504. doi:10.1121/1.394045

Byrne, D., & Cotton, S. (1988). Evaluation of the National Acoustic Laboratories' New Hearing Aid Selection Procedure. *Journal of speech and hearing research, 31*(2), 178-186.

Caposecco, A., Hickson, L., & Meyer, C. (2014). Hearing aid user guides: Suitability for older adults. *International Journal of Audiology, 53*(sup1), S43-S51.

Ching, T. Y. C., Scollie, S. D., Dillon, H., & Seewald, R. (2010). A cross-over, double-blind comparison of the NAL-NL1 and the DSL v4.1 prescriptions for children with mild to

moderately severe hearing loss. *International Journal of Audiology, 49*(S1), S4-S15. doi:10.3109/14992020903148020

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Collins-Thompson, K. (2013). Computational assessment of text readability: A survey of current and future research. *ITL: International Journal of Applied Linguistics, 165*(2), 97-135. doi:10.1075/itl.165.2.01col

Cox, R. M. (2003). Assessment of subjective outcome of hearing aid fitting: Getting the client's point of view. *International Journal of Audiology, 42*(S1), 90-96. doi:10.3109/14992020309074629

Cox, R. M. (2005). Choosing a self-report measure for hearing aid fitting outcomes. *Semin Hear, 26*(3), 149-156. doi:10.1055/s-2005-916378

Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes, 54*(5-6), 340. doi:10.1080/0163853X.2017.1296264

D'Alessandro, D. M., Kingsley, P., & Johnson-West, J. (2001). The readability of pediatric patient education materials on the World Wide Web. *Archives of Pediatrics & Adolescent Medicine, 155*(7), 807-812.

Dahl, T. H. (2002). International classification of functioning, disability and health: An introduction and discussion of its potential impact on rehabilitation services and research. *Journal of Rehabilitation Medicine, 34*(5), 201-204.

Danermark, B., Cieza, A., Gangé, J.-P., Gimigliano, F., Granberg, S., Hickson, L., . . . Swanepoel, D. (2010). International classification of functioning, disability, and health core sets for hearing loss: A discussion paper and invitation. *International Journal of Audiology, 49*(4), 256-262. doi:10.3109/14992020903410110

Davis, T. C., Berkel, H. J., Arnold, C. L., Nandy, I., Jackson, R. H., & Murphy, P. W. (1998). Intervention to increase mammography utilization in a public hospital. *Journal of General Internal Medicine, 13*(4), 230-233. doi:10.1046/j.1525-1497.1998.00072.x

Dean, G., Orford, A., Staines, R., McGee, A., & Smith, K. J. (2017). Psychosocial well-being and health-related quality of life in a UK population with Usher syndrome. *BMJ, 7*(1). doi:10.1136/bmjopen-2016-013261

Doak, C., Doak, L., & Root, J. H. (1996). Teaching patients with low literacy skills. *American Journal of Nursing, 96*(12), 16.

Donald, A. J., & Kelly-Campbell, R. J. (2016). Pediatric audiology report: Assessment and revision of an audiology report written to parents of children with hearing impairment. *Journal of Speech, Language, and Hearing Research, 59*(2), 359-372.

Douglas, A., & Kelly-Campbell, R. J. (2018). Readability of patient-reported outcome measures in adult audiologic rehabilitation. *American Journal of Audiology, 27*(2), 208-218. doi:10.1044/2018_AJA-17-0095

DuBay, W. H. (2004). The principles of readability. *Online Submission.* Retrieved from https://eric.ed.gov/?id=ED490073

El-Daly, I., Ibraheim, H., Rajakulendran, K., Culpan, P., & Bates, P. (2016). Are patient-reported outcome measures in orthopaedics easily read by patients? *Clinical Orthopaedics and Related Research, 474*(1), 246-255. doi:10.1007/s11999-015-4595-0

Engdahl, B., Tambs, K., & Hoffman, H. J. (2013). Otoacoustic emissions, pure-tone audiometry, and self-reported hearing. *International Journal of Audiology, 52*(2), 74-82. doi:10.3109/14992027.2012.733423

Erdman, S. (1993). Self-assessment in audiology: The clinical rationale. *Seminars in Hearing, 14*(4), 303-313. doi:10.1055/s-0028-1085129

Fitzsimmons, P., Michael, B., Hulley, J., & Scott, G. (2010). A readability assessment of online Parkinson's disease information. *The Journal of the Royal College of Physicians of Edinburgh, 40*(4), 292-296.

Friedman, D. B., & Hoffman-Goetz, L. (2006). A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Education & Behavior, 33*(3), 352-373.

Gilligan, J., & Weinstein, B. E. (2014). Health literacy and patient-centered care in audiology: Implications for adult aural rehabilitation. *Journal of Communication Disorders, Deaf Studies & Hearing Aids, 2*, 110.

Greenberg, D. (2001). A critical look at health literacy. *Adult Basic Education, 11*(2), 67;79;.

Greywoode, J., Bluman, E., Spiegel, J., & Boon, M. (2009). Readability analysis of patient information on the American Academy of Otolaryngology–Head and Neck Surgery website. *Otolaryngology - Head and Neck Surgery, 141*(5), 555-558. doi:10.1016/j.otohns.2009.08.004

Hansberry, D. R., Ramchand, T., Patel, S., Kraus, C., Jung, J., Agarwal, N., . . . Baker, S. R. (2014). Are we failing to communicate? Internet-based patient education materials and radiation safety. *European Journal of Radiology, 83*(9), 1698-1702. doi:10.1016/j.ejrad.2014.04.013

Hawkins, D. B. (2005). Effectiveness of counseling-based adult group aural rehabilitation programs: A systematic review of the evidence. *Journal of the American Academy of Audiology, 16*(7), 485-493. doi:10.3766/jaaa.16.7.8

Hawkins, D. B., Montgomery, A. A., Prosek, R. A., & Walden, B. E. (1987). Examination of two issues concerning functional gain measurements. *Journal of Speech and Hearing Disorders, 52*(1), 56-63. doi:10.1044/jshd.5201.56

Hickson, L., Meyer, C., Lovelock, K., Lampert, M., & Khan, A. (2014). Factors associated with success with hearing aids in older adults. *International Journal of Audiology, 53*(S1), S18-S27. doi:10.3109/14992027.2013.860488

Hunter, L. L., & Sanford, C. A. (2014). Tympanometry and wideband acoustic immittance. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (7th ed., pp. 137-163). Philadelphia, PA: Wolters Kluwer Health.

Hyde, M. L. (2000). Reasonable psychometric standards for self-report outcome measures in audiological rehabilitation. *Ear and Hearing, 21*(Suppl), 24-36S. doi:10.1097/00003446-200008001-00005

Ingo, E., Brännström, J., Andersson, G., Lunner, T., Laplante-Lévesque, A., Lund, U., . . . Lunds, u. (2017). Stages of change in audiology: Comparison of three self-assessment measures. *International Journal of Audiology, 56*(7), 516.

Ingvalson, E. M., Lee, B., Fiebig, P., & Wong, P. C. M. (2013). The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients. *Journal of Speech, Language, and Hearing Research, 56*(1), 81-88. doi:10.1044/1092-4388(2012/11-0291)

Kagitcibasi, C., Goksen, F., & Gulgoz, S. (2005). Functional adult literacy and empowerment of women: Impact of a functional literacy program in turkey. *Journal of Adolescent & Adult Literacy, 48*(6), 472-489. doi:10.1598/JAAL.48.6.3

Katz, J. (2014). A brief introduction to clinical audiology and this handbook. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of clinical audiology* (pp. 3-8). Philadelphia, PA: Wolters Kluwer Health.

Kelly-Campbell, R. J., Atcherson, S. R., Zimmerman, K. R., & Zraick, R. I. (2012). Readability of audiologic self-report assessment tools. *Journal of the Academy of Rehabilitative Audiology, 45*, 63-73.

Kelly-Campbell, R. J., & McMillan, A. (2015). The relationship between hearing aid self-efficacy and hearing aid satisfaction. *American Journal of Audiology, 24*(4), 529-535. doi:10.1044/2015_AJA-15-0028

Kickbusch, I. (1997). Think health: What makes the difference? *Health Promotion International, 12*(4), 265-272.

Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* Retrieved from https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

Kong, K., & Hu, A. (2015). Readability assessment of online tracheostomy care resources. *Otolaryngology–Head and Neck Surgery, 152*(2), 272-278. doi:10.1177/0194599814560338

Kreisman, B. M., Smart, J. L., & John, A. B. (2014). Diagnostic audiology. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of clinical audiology* (7th ed., pp. 119-134). Philadelphia, PA: Wolters Kluwer Health.

Matthews, T. L., & Sewell, J. C. (2002). *State official's guide to health literacy*: Council of State Governments.

McArdle, R., & Hnath-Chisolm, T. (2014). Speech audiometry. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of clinical audiology* (7th ed., pp. 61-76). Philadelphia, PA: Wolters Kluwer Health.

McMullan, A., Kelly-Campbell, R. J., & Wise, K. (2018). Improving hearing aid self-efficacy and utility through revising a hearing aid user guide: A pilot study. *American Journal of Audiology, 27*(1), 45-56. doi:10.1044/2017_AJA-17-0035

Moore, B. C. J., Alcántara, J. I., & Marriage, J. (2001). Comparison of three procedures for initial fitting of compression hearing aids. I. Experienced users, fitted bilaterally. *British journal of audiology, 35*(6), 339-353. doi:10.1080/00305364.2001.11745252

Murray, T., Hagey, J., Willms, D., Shillington, R., & Desjardins, R. (2008). Health literacy in Canada: A healthy understanding. *UCLA Previously Published Works.* Retrieved from https://escholarship.org/uc/item/890661nm

Noble, W. (2008). Auditory reality and self-assessment of hearing. *Trends in Amplification, 12*(2), 113-120. doi:10.1177/1084713808316172

Noble, W. (2013). *Self-assessment of hearing* (2nd ed.). San Diego, CA: Plural Publishing.

Northern, J., & Beyer, C. (1999). Reducing hearing aid returns through patient education. *Audiology Today, 11*(2), 10-11.

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine, 67*(12), 2072-2078.

Paasche-Orlow, M. K., Taylor, H. A., & Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readability. *New England journal of medicine, 348*(8), 721-726.

Parker, R. (2000). Health literacy: A challenge for American patients and their health care providers. *Health Promotion International, 15*(4), 277-283. doi:10.1093/heapro/15.4.277

Pires, C., Cavaco, A., & Vigário, M. (2017). Towards the definition of linguistic metrics for evaluating text readability. *Journal of Quantitative Linguistics, 24*(4), 319-349. doi:10.1080/09296174.2017.1311448

Purcell-Gates, V., Degener, S. C., Jacobson, E., & Soler, M. (2002). Impact of authentic adult literacy instruction on adult literacy practices. *Reading Research Quarterly, 37*(1), 70-92. doi:10.1598/RRQ.37.1.3

Quigley, B. A., Coady, M., Gregoire, H., Folinsbee, S., & Kraglund-Gauthier, W. (2009). "More universal for some than others": Canada's health care system and the role of adult education. *New Directions for Adult and Continuing Education*(124), 49.

Ratzan, S. C. (2001). Health literacy: Communication for the public good. *Health Promotion International, 16*(2), 207-214. doi:10.1093/heapro/16.2.207

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: roc area, cohen's d, and r. *Law and Human Behavior, 29*(5), 615-620. doi:10.1007/s10979-005-6832-7

Robinette, M. S., & Glattke, T. J. (2007). *Otoacoustic emissions: Clinical applications* (3rd ed.). New York: Thieme.

Rudd, R. E. (2004). Adult education and public health partner to address health literacy needs (Vol. 15, pp. 7-9). Los Angeles, CA: SAGE Publications.

Schecter, S. R., & Lynch, J. (2011). Health learning and adult education: In search of a theory of practice. *Adult Education Quarterly, 61*(3), 207-224. doi:10.1177/0741713610380438

Schlauch, R. S., & Nelson, P. (2014). Puretone evaluation. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of clinical audiology* (7th ed., pp. 29-48). Philadelphia, PA: Wolters Kluwer Health.

Selden, C. R., Zorn, M., Ratzan, S. C., & Parker, R. M. (2000). Health literacy. *Current bibliographies in medicine, 1*, 2000.

Sindhusake, D., Mitchell, P., Smith, W., Golding, M., Newall, P., Hartley, D., & Rubin, G. (2001). Validation of self-reported hearing loss: The Blue Mountains hearing study. *International journal of epidemiology, 30*(6), 1371-1378. doi:10.1093/ije/30.6.1371

Smith, S. L., & West, R. L. (2006). The application of self-efficacy principles to audiologic rehabilitation: A tutorial. *American Journal of Audiology, 15*(1), 46-56. doi:10.1044/1059-0889(2006/006)

Stevens, G., Flaxman, S., Brunskill, E., Mascarenhas, M., Mathers, C. D., & Finucane, M. (2013). Global and regional hearing impairment prevalence: An analysis of 42 studies in 29 countries. *European Journal of Public Health, 23*(1), 146-152. doi:10.1093/eurpub/ckr176

Stucki, G., Cieza, A., Ewert, T., Kostanjsek, N., Chatterji, S., & ÜstÜn, T. B. (2002). Application of the International Classification of Functioning, Disability and Health (ICF) in clinical practice. *Disability & Rehabilitation, 24*(5), 281-282. doi:10.1080/09638280110105222

Stucki, G., & Grimby, G. (2004). Foreword. *Journal of Rehabilitation Medicine, 36*, 5-6. doi:10.1080/16501960410022300

Svarstad, B. L., Bultman, D. C., Mount, J. K., & Tabak, E. R. (2003). Evaluation of written prescription information provided in community pharmacies: A study in eight states. *Journal of the American Pharmacists Association, 43*(3), 383-393.

Sweetow, R., & Palmer, C. V. (2005). Efficacy of individual auditory training in adults: A systematic review of the evidence. *Journal of the American Academy of Audiology, 16*(7), 494-504. doi:10.3766/jaaa.16.7.9

Valente, M., & Valente, M. (2014). Hearing aid fitting for adults: Selection, fitting, verification, and validation. In J. Katz, M. Chasin, K. M. English, L. J. Hood, & K. L.

Tillery (Eds.), *Handbook of clinical audiology* (7th ed., pp. 777-804). Philadelphia, PA: Wolters Kluwer Health.

Vestergaard, M. D. (2006). Self-report outcome in new hearing-aid users: Longitudinal trends and relationships between subjective measures of benefit and satisfaction. *International Journal of Audiology, 45*(7), 382-392. doi:10.1080/14992020600690977

Wang, L.-W., Miller, M. J., Schmitt, M. R., & Wen, F. K. (2013). Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Research in Social and Administrative Pharmacy, 9*(5), 503-516. doi:10.1016/j.sapharm.2012.05.009

Weiss, B. D. (2007). *Health literacy and patient safety: Help patients understand. Manual for clinicians*. Chicago, IL: American Medical Association Foundation.

Welling, D. R., & Ukstins, C. A. (2015). *Fundamentals of audiology for the speech-language pathologist.*

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 50*(4), 844-856. doi:10.1044/1092-4388(2007/059)

Wolf, M. S., Feinglass, J., Thompson, J., & Baker, D. W. (2010). In search of 'low health literacy': Threshold vs. gradient effect of literacy on health status and mortality. *Social Science & Medicine, 70*(9), 1335-1341. doi:10.1016/j.socscimed.2009.12.013

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. (8024715872). Geneva: World Health Organization.

World Health Organization. (2018). Deafness and hearing loss: Key facts. Retrieved from http://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss

World Health Organization. (n.d.). Grades of hearing impairment. Retrieved from http://www.who.int/pbd/deafness/hearing_impairment_grades/en/

Yin, H. S., Forbis, S. G., & Dreyer, B. P. (2007). Health literacy and pediatric health. *Current problems in pediatric and adolescent health care, 7*(37), 258-286.

Zaballos, M., Plasencia, D., González, M., De Miguel, A., & Macías, Á. (2016). Air traffic controllers' long-term speech-in-noise training effects: A control group study. *Noise and Health, 18*(85), 376-381. doi:10.4103/1463-1741.195804

Zakaluk, B. L., & Samuels, S. J. (1988). *Readability: Its past, present, and future*.

Zraick, R. I., Atcherson, S. R., & Brown, A. M. (2012). Readability of patient-reported outcome questionnaires for use with persons who stutter. *Journal of Fluency Disorders, 37*(1), 20-24. doi:10.1016/j.jfludis.2011.10.004