LEARNING OUTCOMES OF

SPEECH AUDIOMETRY VIRTUAL PATIENT SIMULATOR USE

FOR EXPERT AND NOVICE AUDIOLOGY STUDENTS

A thesis submitted in partial fulfillment of the requirements for the

Degree of

Master of Audiology

in the University of Canterbury

By GERARD WILLIAM

Department of Communication Disorders

University of Canterbury

2013
Abstract

Rationale:

Audiology student training in New Zealand faces many difficulties with a limited number of qualified instructors and suitable external placements. With a continued shortage of audiologists in New Zealand, new methods of training need to be introduced and implemented. One solution is through the use of realistic, computer-based virtual patient simulators (VPS). HIT Lab New Zealand in conjunction with the University of Canterbury has designed a VPS for New Zealand audiology students. A speech audiology component is to be developed based on best practice recommendations, and needs to be validated.

Method:

Two studies, one with 18 Master of Audiology (“expert”) and another with 18 (“novice”) undergraduate students, were evenly divided into simulator and non-simulator user groups. Simulator users had to complete 5 virtual patient cases in addition to the non-simulator users’ requirement to refer to provided lecture notes and speech audiometry protocols. Novice students were assessed on declarative, procedural and retained knowledge of speech audiometry; expert students were additionally assessed on training transfer. The intervention period was set at two weeks, and the retention assessment at four weeks post-intervention.
Results:

Expert students who used the simulator significantly improved their training transfer skills. No significant differences were found between and within groups for declarative knowledge and procedural knowledge. Training transfer and procedural knowledge were retained for both groups, but only non-simulator users retained declarative knowledge.

Novice students who used the simulator significantly increased their declarative knowledge. Both groups’ procedural knowledge significantly regressed post-intervention. Declarative and procedural knowledge were retained for both groups.

Implications:

Simulator use appears to accelerate learning outcomes otherwise achievable through traditional learning methods, and does depend on the users’ existing knowledge base. Regular use may be necessary to retain desired learning outcomes. Improvements (e.g., more detailed feedback systems) are to be incorporated into the simulator, and sole reliance on the simulator for learning is not recommended. Future research into more holistic aspects of virtual patient use within the field of audiology and allied health care is warranted.
Acknowledgements

This research was funded by grants through HIT Lab New Zealand (TEC-ESI Grant) and GN ReSound (ReSearch Grant). I also acknowledge the support of the University of Canterbury. I am grateful to the Department of Communication Disorders who has supported this research and the HIT Lab programmers who have been generous with their feedback and support.

Personally, this thesis would be impossible without the help of Dr Catherine Moran and Mr Jonny Grady, who have always made the time to assist me and handle every query or difficulty. Thank you. To my fellow researchers in the field, Alexandre Heitz and Lynda Guard, you have both been amazing and have gone out of your way to ensure the smooth completion of this project. Dion Jurac, Alice Bowden and Jakob Zillner also need to be commended for the time they have spent on the implementation and the design of this research respectively. My fellow audiology students and future colleagues have participated wonderfully in this study, although they were free not to: I appreciate the efforts every one of you have made.

I would especially like to thank my family for their continued and positive support over the course of my studies. I would like to mention Marama and Scott Humphreys for their proofreading support and assistance in a new land. In particular, my wife Jane who has patiently supported me in every move, and continually kept me motivated at every step. I love you. Finally, and most importantly, I wish to thank Jesus Christ, through whom all things are possible.

God bless you all. Gerard William (A.M.D.G.)
# Table of Contents

**Introduction:** ........................................................................................................... 1

**This study:** ............................................................................................................... 4

**Literature Review** ..................................................................................................... 5

**Speech Audiometry in New Zealand** ........................................................................ 5

- The AB-CVC meaningful word lists ......................................................................... 8
- Criticisms of Speech Audiometry ............................................................................. 11
- Learning Speech Audiometry .................................................................................. 13

**Use of Virtual Patient Simulators (VPS) for Learning** ............................................ 15

- Theoretical models of virtual patient simulator use ................................................. 19
- Learning outcomes of virtual patient simulator use ............................................... 23
- Best Practice design principles for Virtual Patient Simulators ................................ 28

**Summary and research questions** ........................................................................... 31

**Hypotheses** ............................................................................................................. 32

**Development of simulator and instruments** ............................................................. 34

- Speech Audiology Simulator Development ............................................................. 34

**Development of assessment measures of learning outcomes:** ............................... 37

- Training Transfer ..................................................................................................... 37
- Procedural Knowledge ............................................................................................ 39
- Declarative Knowledge ............................................................................................ 41
- Retained Knowledge ............................................................................................... 42
- Validity of measures ............................................................................................... 42
- Reliability estimates ............................................................................................... 43
- Pilot testing .............................................................................................................. 43
Method ........................................................................................................................................ 44

Study 1: Advanced Audiology Students (AAS) ........................................................................... 44

  Participants.................................................................................................................................. 44
  Materials ...................................................................................................................................... 44
  Equipment .................................................................................................................................... 45
  General procedure .......................................................................................................................... 46

Study 2: Novice audiology students (NAS) .................................................................................... 48

  Participants.................................................................................................................................. 48
  Materials ...................................................................................................................................... 48
  General procedure .......................................................................................................................... 49

Research Design ............................................................................................................................ 49

Results ............................................................................................................................................ 52

  Descriptive Statistics .................................................................................................................... 52

  Summary of Results ...................................................................................................................... 53

Advanced Audiology students (AAS) ............................................................................................ 54

  Training Transfer for AAS ........................................................................................................... 55
  Declarative Knowledge for AAS .................................................................................................... 57
  Procedural Knowledge for AAS .................................................................................................... 59
  Knowledge Retention for AAS ....................................................................................................... 60

Novice Audiology students (NAS) .................................................................................................. 61

  Declarative Knowledge for NAS .................................................................................................. 62
  Procedural Knowledge for NAS .................................................................................................... 63
  Knowledge Retention for NAS ....................................................................................................... 65

Reliability reports ............................................................................................................................. 66

  Training transfer ........................................................................................................................... 66
Declarative knowledge ........................................................................................................ 66
Procedural knowledge .......................................................................................................... 67
Other findings ...................................................................................................................... 67
Simulator Users .................................................................................................................. 67

Discussion ........................................................................................................................ 69

Advanced Audiology Students (AAS) .................................................................................. 69
Novice Audiology Students (NAS) ...................................................................................... 71
Other Findings: Simulator Users ....................................................................................... 73

Limitations .......................................................................................................................... 75

Theoretical Implications ..................................................................................................... 79
Practical implications .......................................................................................................... 83
Future research ...................................................................................................................... 86

Conclusions ......................................................................................................................... 87

References .......................................................................................................................... 90

Appendices .......................................................................................................................... 95

Appendix 1: Training Transfer Marking Schedule for Standardised Patients ................. 95
Appendix 2A: Procedural knowledge questions at Time 1 ................................................. 97
Appendix 2B: Procedural knowledge questions at Time 2 ................................................ 103
Appendix 2C: Procedural knowledge questions at Time 3 ................................................ 109
Appendix 3: Declarative knowledge questions .................................................................... 115
**Introduction:**

There is a continued shortage of suitably trained and qualified audiologists in New Zealand. At the time of writing, audiologists were still on the Department of Labour’s long-term skill shortage list (Immigration New Zealand, 2012). Working as an Audiologist in New Zealand requires both a post-graduate qualification in Audiology, as well as completion of a Certificate of Clinical Competence (CCC) administered by the New Zealand Audiological Society (NZAS). Completing this CCC process includes assessments of the clinical audiologist’s skills in performing the audiological test battery to the NZAS’ standards and protocols. The NZAS audiological test battery for an adult diagnostic assessment minimally includes case history taking, pure-tone audiometry, speech audiometry, tympanometry and acoustic reflex testing. Based on the integrated results of these components, a suitable diagnosis of the configuration, magnitude and origin of the hearing loss can usually be made (McArdle & Hnath-Chisolm, 2009). The NZAS is responsible for setting “best-practice” clinical protocols for New Zealand audiologists, which are adapted for use by both public (e.g. hospitals) and private institutions (e.g. independent clinics, universities). Graduate training of audiologists in New Zealand according to the NZAS protocols is primarily undertaken at two universities: the University of Auckland and the University of Canterbury.

Training in these institutions have followed a traditional model in recent years with a mixture of course papers, clinical supervision and a research based Master’s thesis. Students are expected to be proficient at clinical skills early in their studies. Integrating and consolidating the various components of the diagnostic test battery quickly and effectively can be demanding for the student. In a busy clinic, performing any of these tasks poorly or sub-standard may mean that
the student is not asked to directly perform these tasks again for that day. This can be particularly challenging for different types of students based on their learning styles. Some who prefer a less stressful approach may choose to practice in their spare time, while others may find learning on the job adequate for advancing their clinical skills. Mastering these components to NZAS standards usually requires the trainee audiologist to practice these aspects either in their spare time with volunteers, simulated patients, or under direct supervision on clinical placements (Wilson, Hughes, Sher, & Laplante-Levesque, 2010).

Some parts of the audiological test battery require more specific training to achieve competence. In particular, the speech audiometry procedure can appear to be complicated for trainee audiologists (New Zealand Audiological Society, 2007). Speech audiometry is an essential part of the adult audiological diagnostic assessment (McArdle & Hnath-Chisolm, 2009; Sincock, 2008). Learning the procedure and understanding speech audiometry successfully and efficiently requires practice, instruction and feedback (Wilson et al., 2010). This process is time intensive for both clinical educators and students, especially in the early stages where more intense and focused individual instruction is usually required. Audiology student training in New Zealand faces many difficulties with a limited number of qualified instructors and suitable external placements providing adequate hands-on experiences, which is not dissimilar to challenges faced by audiology and other health professions overseas (Wilson et al., 2010; Issenberg, McGaghie, Petrusa, Gordon & Scalese, 2005; Deladisma, Johnsen, Raij, Rossen, Kotranza, Kalapurakal, Szlam et al., 2008). These factors consequently affect the number of students accepted into audiology programmes at these institutions annually.

Despite extensive review into these problems, novel learning methods are still required. New methods to minimise patient risk and maximise the opportunities for students to learn are
being evaluated in order to train more graduates (Wilson et al., 2010). Novel methods must also ensure that course quality remains at a high standard. Some of these methods have included the use of standardised patients and computer-based simulations. A newer solution, which combines the advantages of both methods, is the use of realistic, computer-based virtual patient simulators (VPS). Virtual patient simulations reduce the need for another person to be present, reduces risks to clients, and if unlimited access is provided, allows the learner to practice and develop skills in their spare time (Cook & Triola, 2009; Wilson et al., 2010; Sitzmann, 2011; Issenberg et al., 2005). However, it is not enough to use previously developed audiology simulators. A dearth of published research currently exists on such simulators. The main commercially available simulators (such as AudSim,1 Otis: The virtual patient,2 and Audiology Clinic3) have their uses, but the main disadvantages of these simulators are that they do not have an interactive interpretation of the client’s response, and do not have a complete test battery developed for New Zealand protocols and stimuli. The NZAS test battery shares similarities with other countries but maintains key differences.

Due to the on-going need for a realistic New Zealand-specific training tool, the increased acceptance of the use of simulators within audiology (Wilson et al., 2010), and the NZAS’ continued preferred use of the AB-CVC word lists (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy, Arlington & Johnstone, 2000), a computer-based virtual patient simulator is being developed for use at the University of Canterbury in conjunction with the HIT Lab New Zealand

1 AudSim – Audiometer Simulation Software. Available from http://www.audsim.com


(Düenser, Heitz & Moran, 2010). Consistent with previous research (Issenberg et al., 2005; Huwendiek, Reichert, Bosse, de Leng, van der Vleuten, Haag, Hoffman & Tönshoff, 2009; Botezatu, Hult, Tessma & Fors, 2010a; 2010b; Deladisma et al., 2008; Cook & Triola, 2009; Sitzmann, 2011), the virtual patient simulator intends to provide unrestricted access to high-fidelity virtual patient cases across various difficulty levels, and students can perform components of the NZAS version of the audiological test battery to hone their clinical skills in a relatively stress-free environment (Wilson et al., 2010).

Testing has been undertaken on the already developed case history and pure-tone audiometry components of the simulator (Howland, 2012). The speech audiometry component of the virtual patient simulator is to be developed and validated. It is the aim of this research to incorporate suitable learning goals from previous research to evaluate the expected utility of the virtual patient speech audiometry simulator amongst current and potential students of audiology.

*This study:*

The literature review that follows will present themes of speech audiometry use in New Zealand clinical settings, before providing a critical insight into the use of virtual patient simulators for learning, particularly within audiology and other allied health professions. Learning outcomes will be identified and evaluated through the derived hypotheses. The methodology for developing the measures and validating the simulator is presented, before the results and discussion provide learned outcomes, implications and future research directives.
Literature Review

Speech Audiometry in New Zealand

Speech audiometry is the component of audiological assessment that assesses an individuals’ ability to respond to speech stimuli in quiet or noisy situations, by completing tasks such as identification or discrimination (McArdle & Hnath-Chisolm, 2009). Speech is fundamental to human communication. Pathologies that affect speech understanding can seriously affect the ability to communicate effectively. Hearing aid prescriptions are usually based on formulae to maximise speech intelligibility. Speech stimuli are used for various other tests in audiology, such as for determining most comfortable and loudness discomfort levels, or verifying hearing aids through real-ear measurements (McArdle & Hnath-Chisolm, 2009; Punch, Rakerd & Joseph, 2004). Thus, there is a strong need to perform speech audiometry testing (Boothroyd, 2008).

The integrity of the auditory pathway may be compromised if diagnostic assessments exclude speech audiometry (McArdle & Hnath-Chisolm, 2009; Martin, 2009). Complex sounds such as speech stimuli require deeper neural processing encompassing more areas of the auditory pathway, and subtle deficits in processing abilities can cause greater impairments in tasks such as speech discrimination. These deficits may not be evident with simple pure-tone stimuli, as patients with normal or near-normal audiograms may have greater difficulty when tested with speech audiometry. These deficits can be observed in cases of vestibular schwanommas, Auditory Processing Disorders (APD) or some expressions of Auditory Neuropathy Spectrum Disorder (ANSD), where the sound may be audible but there is a distinct lack of speech intelligibility (McArdle & Hnath-Chisholm, 2009). To a lesser extent, sensorineural hearing
losses may also sometimes present sub-optimal scores in speech tests at raised intensity levels (Boothroyd & Nittrouer, 1988; Boothroyd, 2008). Therefore, knowing how an individual with a hearing loss performs with speech stimuli has substantial clinical utility. Firstly, it demonstrates how an individual performs with speech stimuli at normal conversational levels in a quiet setting. The face validity provided to the client can also be useful in some situations. For example, a patient scoring 80% at a 60 dB HL input level may be counseled that they can only successfully discriminate about 80% of the words at a normal conversational level in a quiet setting (Boothroyd, 2008; McArdle & Hnath-Chisholm, 2009). Secondly, it provides an indicator as to whether amplification is potentially beneficial as determined by better performance at raised intensity levels. Finally, speech audiometry can provide a cross-check tool of internal consistency with pure-tone audiometry (Boothroyd, 2008). Consistency checks are pertinent in cases of functional hearing losses where speech audiometry is likely to significantly outperform pure-tone audiometry. Proceeding to obtain a Performance-Intensity function means that their optimal listening range, expected improvement with increasing audibility, and the consistency of their performance with their audiogram can be ascertained (Boothroyd, 2008).

There are many versions of speech audiometry tests which incorporate words, sentences or nonsense words. For example, the Central Institute for the Deaf in the United States of America developed a commonly used everyday sentence test or Auditory Test (CID W-22). Harvard Psychoacoustics Laboratory also produced 24 word lists of 50 phonetically balanced monosyllabic words (PAL PB-50 Word lists) used for testing adults. Each word is preceded by a carrier phrase, and scoring is based on the entire word. The use of this closed-set testing of spondaic words presented monaurally is to obtain a speech reception threshold (SRT). The SRT corresponds to the listener’s audiogram; however, there is limited utility beyond consistency
purposes. Northwestern University also produced Auditory Tests (NU-4; NU-6) that use a Consonant-Nucleus-Consonant structure in male and female voice recordings. They also developed the NU-CHIPS (Children’s Perception of Speech) consonant and nonsense syllable test frequently used for testing young and school-aged children.

In New Zealand, speech audiometry testing can be either with words or sentences, performed primarily in quiet, but sometimes in noise. The two main tests for adult testing are the Arthur Boothroyd (AB) meaningful CVC (Consonant-Vowel-Consonant) words lists in quiet (Boothroyd, 1968; 2008; Boothroyd & Nittouer, 1988; Purdy et al., 2000) and the Hearing in Noise Test (HINT). The AB-CVC words list is used as an open set test for testing adults in quiet, and forms the main basis of speech audiometry testing in New Zealand. Since the AB-CVC word lists are presented at varying intensity levels for each ear, a Performance-Intensity function can be obtained relatively quickly for each ear, providing a quick insight into the speech discrimination function of each ear to speech stimuli in quiet (Boothroyd, 2008). The HINT sentences are predominantly used for cochlear implant candidacy assessment where it is used in auditory alone and auditory-visual conditions. Other tests that are used in clinical situations to varying degrees include QuickSin, where test sentences are presented at a supra-threshold level in degrading signal-to-noise ratios. This is particularly useful in demonstrating to clients their difficulty understanding speech in noise, and is not as closely correlated to their audiograms. The common theme in these tests is that they attempt to ascertain if an individual is able to identify certain components of speech (all tests), process speech information in quiet (AB-CVC) or in noise (HINT, QuickSin).
The AB-CVC meaningful word lists

Arthur Boothroyd first developed the AB-CVC word lists in 1968 (see Boothroyd, 1968; Martin, 1987). Originally recorded with a Northern English accent, it was later recorded in other accents including the Australian accent (Martin, 1987). The word lists were later updated to include terms in more common usage (Boothroyd and Nittrouer, 1988). A New Zealand recording of the twelve wordlists (120 words in total) with a native New Zealand male speaker was produced and normalised by Purdy (Purdy et al., 2000). Its relative ease to implement and use of a New Zealand accent means that the AB-CVC word list has remained the gold standard for speech audiometry testing in New Zealand. Testing involves presenting each list of ten phonetically balanced isophonemic words to both ears, one ear at a time. Each word list is presented at a given intensity level; and each word is preceded by a carrier phrase (“Say”) and consists of a consonant, vowel and another consonant. Presentation levels to the tested ear and required masking for the contralateral ear is directly related to the pure-tone audiogram, and incorporates the magnitude, origin and nature of the hearing loss (McArdle & Hnath-Chisholm, 2009; Yacullo, 2009). If there is a chance that the non-test ear might be able to detect the speech signal, then it needs to be masked with speech-weighted broadband noise (Yacullo, 2009). The client is asked to repeat each word after the speaker, and each word is scored phonetically (Boothroyd, 2008; Purdy et al., 2000). By scoring the constituent phonemes of each word, there is increased reliability of the resultant function, making it easy to obtain scores from different levels from multiple word lists relatively quickly. It removes tester or participant bias by discounting the listener’s vocabulary knowledge and assesses them purely on their auditory resolution and ability to articulate the sounds that they have heard (Boothroyd, 2008).
The total score for the ten words in each list is plotted. Different intensity levels are used for each ear to obtain a PI function (see figure below). The pattern of results for each ear follows a psychometric function depending on the nature of the auditory system. This cubed exponential function can vary; and the distinct shape and position of each curve indicates the speech understanding capability of each ear (McArdle & Hnath-Chisolm, 2009; Boothroyd, 2008). The PI function is transformed by the nature of the underlying pathologies. For example, a conductive loss can shift the a PI-function to the right by the amount of average hearing loss, but a sensorineural loss can result in sub-optimal PI functions even at raised intensity levels.

The PI function has two main components: the Performance Intensity Maximum (PI-Max) score (mentioned earlier) and a Half-peak level (HPL). The PI-Max is the optimal score a patient may obtain across all intensity levels. Speech discrimination for normal hearing is represented by maximum PI scores (PI-Max) of approximately 100% between 30 to 40 dB HL (McArdle & Hnath-Chisolm, 2009). For some pathologies such as vestibular schwannomas, higher intensity stimuli (e.g. At 90 dB HL) may produce worse scores than at lower intensity levels. This decrease in PI-Max score with increasing intensity is called Rollover. The Rollover coefficient was a useful tool in early identification of retrocochlear hearing losses prior to the more affordable proliferation of Magnetic Resonance Imaging (MRI) (Martin, 1987). On the other hand, better than expected scores may reveal pseudohyperacusis, malingering or a
functional hearing loss since it is difficult to fake speech audiometry performance (Martin, 2009).

The PI-function may be used as a predictor of hearing aid benefit (Boothroyd, 2008). If raised intensity levels can show improved speech discrimination, then it is logical to expect that amplification may also be expected to do so. This is not without its caveats, since hearing aid adoption remains a multi-faceted and complex process with many exceptions. However, the use of speech audiometry may still be relatively effective at identifying those with expected hearing aid benefit from those of whom no reasonable benefit is expected.

*Figure 1: A diagram of a Performance Intensity (PI) function for a flat mild-to-moderately-severe sensorineural hearing loss. Source: Boothroyd (2008, p.484).*

---

Fig. 7. Pure-tone audiogram and binaural PI functions, via headphones and aided, for an adult with a moderate sensorineural hearing loss. Note that PRT stands for Phoneme Recognition Threshold and is the level in dB, re normal, at which the listener scores 50% of his maximum. PRT is similar to the more familiar SRT.
Criticisms of Speech Audiometry

Criticism has been leveled at the PI function for its lack of sensitivity (McArdle & Hnath-Chisolm, 2009; Martin, 1987), but proponents for it argue for its utility and underuse (Boothroyd, 2008). Despite phonetic balancing, higher frequency losses can result in bigger shifts in the PI function. Presentation levels are limited to around 90 dB HL. Beyond this intensity level, the chosen transducers may distort and clip, which may negatively skew performance scores.

Speech audiometry is not very sensitive for detecting peripheral pathologies. Presentation levels are limited and speech stimuli are limited by current transducers ability to present at levels above 90 dB HL. This means that severe or worse losses cannot be tested through current speech audiometry techniques. Fortunately, these can be identified through the other components of the diagnostic battery. However, the use of speech audiometry as part of the speech battery increases its diagnostic value, and when the goals of rehabilitation to improve speech understanding are considered, its importance is still emphasised. For retrocochlear losses, Magnetic Resonance Imaging (MRI) are the accepted gold standard approach to identifying associated pathologies. As MRI costs continue to fall, the use of speech audiometry to sensitively identify these pathologies will continue to decrease. However, speech discrimination testing remains a more affordable and practical early identifier to determine if a patient needs to undergo an MRI.

Some critics insist that hearing aid benefit cannot be determined through speech audiometry (McArdle & Hnath-Chisolm, 2009). Speech testing in quiet cannot be used to extrapolate performance for understanding speech in the real world. However, it still provides an indication whether the problem of audibility can alleviate discrimination concerns (Boothroyd,
If hearing aid benefit is the end outcome, it is always crucial to consider that hearing aid use and adoption is a complex, multifaceted decision (Tyre-Murray, 2009).

Two recommendations are commonly proposed to overcome current limitations in speech audiometry: using a speech in noise test, or increasing the difficulty by using open- instead of a closed-set test (McArdle & Hnath-Chisolm, 2009). Tests such as HINT sentences are sometimes preferred in addition to speech-in-quiet tests. If the main difficulty identified by the client is in “noisy situations”, then testing should include both speech stimuli in quiet and in noise settings, to test for the reasons identified above, as well as to obtain more reliable information about changes in the minimum signal-to-noise ratio (SNR) required for the client. Tests such as QuickSin when used have the advantages of immediate face validity for the client, as well as an indication of the client’s ability to identify the target speech signal over any presented noise.

The second solution is the use of open-set testing instead of closed-set testing. In closed-set lists, the client provides a response from a list of familiar words, or is provided a context for the word. To mimic real-life situations, open-ended speech stimuli with no prior clues or suggestions should be used. Open-set testing also ensures that random guesses or chance is greatly minimised or eliminated (Sincock, 2008; Tyre-Murray 2009). The AB-CVC word lists are an example of open-set testing (Purdy et al., 2000; Boothroyd and Nittroeur, 1988).

Despite these criticisms, it is expected that for the foreseeable future, the preferred speech audiometry tool utilised in New Zealand as determined by the NZAS remains the AB-CVC word list until a new gold standard is established (New Zealand Audiological Society, 2007). Compressed and reverberated versions of the AB word lists are in current use for APD testing, and current research leans towards further development of video based tests, more open-set
sentence discrimination testing and adaptive speech testing (McArdle & Hnath-Chisolm, 2009; Sincock, 2008). As it stands, these are still not in widespread use due to their lack of generalisability and efficiency. Coupled with the general misunderstanding and underuse of the PI Function, greater use of speech audiometry is in fact recommended (Boothroyd, 2008).

**Learning Speech Audiometry**

Nine tasks of speech audiometry using the AB word lists can be summarised as follows:

1. Deciding on which transducer to use
2. Giving instructions and calibrating the system
3. Choosing appropriate presentation levels for the better ear
4. Deciding on the need for masking and level of masking
5. Recording and scoring the results
6. Obtaining a PI-Max and Half-peak level
7. Repeating the test procedure in the other ear
8. Checking consistency with audiogram
9. Feeding the results back to the client

The New Zealand Audiological Society requires trained audiologists to be competent at what they do (New Zealand Audiological Society, 1998). They establish protocols for full membership registration, called the Certificate of Clinical Competency, which practicing audiologists need to obtain to be accredited as full members and to undertake work authorised by Governmental organisations. As a component of the adult test battery, speech audiometry testing requires test protocols. Although intended to be clinically efficient, it is a fairly complicated process that is not easy to learn. Practical complications for the nine tasks include the multi-tasking nature of testing, which includes watching the subject responding, controlling the CD-
player, writing down and scoring responses, and encouraging the client. Technical challenges can include efficiently and accurately deriving optimal presentation levels to obtain a PI-max and a corresponding half-peak level, working out the need for and amount of masking for each word list, and evaluating the consistency of test scores to the pure-tone audiometry. An ideal clinical situation requires this testing to be completed well under fifteen minutes, which is a relatively short testing timeframe in a clinical setting.

Student audiologists can be overwhelmed with the number of concepts and procedures they have to remember and assimilate, especially when considering the numerous other tests and pathologies they are also required to know as part of the audiological battery (Wilson et al., 2010). Being competent and efficient with speech audiometry requires practice and ideally one-to-one supervision. Focused supervision is difficult to provide in busy clinical courses, which consequently reduces the maximum number of students accepted into audiological courses annually (Wilson et al., 2010). Students are currently taught to be competent while learning and interacting with actual clients. Expecting audiology students to both familiarise themselves to laboratory equipment, as well as establish competency for all test procedures quickly, can be difficult and discouraging for students (Wilson et al., 2010). If students do not feel confident, they may find themselves making mistakes and negatively evaluated for it, which may affect their learning and coursework (Wilson et al., 2010). This would in turn affect the quality of the taught audiology programme. Hence, if other methods were available to teach modules of the test battery, such as speech audiometry, then this would be beneficial to audiology students, clinical educators and training institutions.
**Use of Virtual Patient Simulators (VPS) for Learning**

Computer-based simulations are a recent solution has been introduced to overcome the problem of limited learning opportunities. Simulations are able to replace or recreate real experiences, and evoke substantial elements of the real world in a fully interactive manner (Anderson & Warren, 2011). Computer-based simulators provide the advantages of real-life patient testing without its disadvantages. Computer simulations, compared to human standardised patients, do not become embarrassed or stressed, and predictable behaviour can be programmed for a variety of situations and needs (Issenberg, McGaghie, Hart, Mayer, Felner, Petrusa, Waugh et al., 1999). In addition to lower patient and clinician risks (Lieberth & Martin, 2005; Dawson, Cotin, Megalan, Shaffer & Ferrell, 2000; Sitzmann, 2011; Willaert, Aggarwal, Van Herzeele, Cheshire & Vermassen, 2012), there are several added advantages such as the provision of rare cases (Lieberth & Martin, 2005; Dawson et al., 2000; Sitzmann, 2011), the freedom for students to access it as needed (Sitzmann, 2011), and to practice as often as they like (Wilson et al., 2010; Sitzmann, 2011). Users can also freely make mistakes and self-correct. They can be incorporated into curriculum for training and assessment purposes, since they allow for the training of procedures and management of difficult situations within a standardised, reproducible experience for all users (Issenberg et al. 1999). Thus, they can overcome the common problems of inadequate clinical learning opportunities, instruction time constraints, limited patient availability for hands on experiences faced by many allied health professions such as audiology, nursing, and pharmacy, as well as medical and surgical professions (e.g. Issenberg et al., 1999; Dawson & Gould, 2007; Dawson et al., 2000; Kane-Gill & Smithburger, 2011; Wilson et al., 2010; Anderson & Warren, 2011; Wong, 2004; Vyas, Wombwell, Russell & Caliguiri, 2010; Vyas, Ottis & Caliguiri 2011). Overall, it is a flexible and empowering learning solution.
Prior to the proliferation of computer-based technologies, human simulations such as standardised patient (SP) testing has been, and still remains, extremely useful for training purposes (Vyas et al., 2010; Vyas et al., 2011; Issenberg et al., 1999). SP testing allows the student to practice on a real person. Mistakes can be self-corrected without major implications. Since a human person is being used, important interpersonal skills can be practiced and enhanced, and has been shown to transfer well into real clinical situations (Kane-Gill & Smithburger, 2011). Finally, due to its consistency, several students can be assessed on one SP.

SP testing unfortunately has its disadvantages. Case quality is largely dependent on well-written cases by examiners, and reliable and realistic acting from the patient (Issenberg et al., 1999). The cost of hiring good SPs can consequently be high, and practice times are limited to SP availability (Tharpe & Rokuson, 2010). Pathologies for each case are also limited to ones that the actors are able to portray. As mentioned, there are also risks associated with practicing on human patients. Computer simulations on the other hand, are able to overcome some of these disadvantages by providing unlimited practices with little to no risk for each student. Initial cost of purchasing or designing the software may be high (Sitzmann, 2011), but subsequent usage costs can often be less in the long run. In addition, mistakes on the simulator can also be self-corrected without implications. With more computer-based learning required across all aspects of coursework, there is increased computer literacy amongst students of all ages. However, computer-based simulations are also dependent on realism and fidelity to real cases. Unrealistic simulations may affect practice and the development of important interpersonal skills.

The exponential advances in computational power and programming capabilities have ensured that computer simulations of greater detail and fidelity are being produced. Graphical user interfaces which give users greater control and interaction capabilities allow those who were
previously adverse to computers to overcome such barriers. Recent advances in computational power has ensured smooth operating intensive 3D applications can achieve many of the benefits previously limited to 2D simulations (Sitzmann, 2011). Therefore, low-fidelity, two-dimensional interfaces can now be replaced by high fidelity, three-dimensional interactions. The falling costs for producing simulator software can assist with its development and distribution, making it more attractive for training centres (e.g. Audiology schools) to adopt simulation-based training. The proliferation of high fidelity simulations ensure a more realistic and engaging platform for learning. More realistic simulations provide more engaging simulations, which result in more active participation (Sitzmann, 2011). Amidst these changes, computer simulations have quickly evolved from simple simulations to complex virtual patient software.

Computer-simulation can be static or dynamic, and involve realistic 3-Dimensional (3D) images or simple 2-Dimensional (2D) graphic representations (Siemers, Fritzson & Nakhimovski, 2009; Sitzmann, 2011). Simple simulations have the advantages of being easier to use whilst providing the same core content as a 3D simulation (Sitzmann, 2011; Lieberth & Martin, 2005). Further, it can be more easily distributed over a web platform due to its smaller software size, which is crucial for distance learning programmes (Sitzmann, 2011; Lieberth & Martin, 2005). However, this advantage may not be directly applicable to audiological training since a core dimension of audiology is to enhance personal client-clinician interactions.

Game based simulations have also been commonly used for training purposes. Gaming deeply engages the user into the learning process, and when the difficulty level and the player’s skill are appropriately matched, an engaging experience can be provided (Csikszentmihalyi, 1990). Where the end is to train a specific skill, gaming can overcome emotional blockages such as boredom and associations with work, by replacing it with the desire to complete the game
In medically related fields, and particularly Audiology, games may not be deemed as a suitable clinical training tool, since the end goal is still to enhance patient-clinician interactions and clinician reasoning skills (Cook & Triola, 2009). However, engagement principles can still be applied within a simulation context. Hence, a highly stimulating yet engaging platform, which combines the best of gaming within a more formalized training design, should be developed.

The mode of simulation is directly related to its profession of use. Simulator training in the fields of engineering and business have seen success through the use of games (Siemers et al., 2009; Sitzmann, 2011). Within healthcare, anatomical simulators are widely used within surgical training to teach anesthetists and cardiologists (Wong, 2004; Dawson et al., 2000) and for dental training (Cook & Triola, 2009; Vyas et al., 2011; Bray, Schwartz, Odegard, Hammer & Seybert, 2011). Anatomical and virtual patient simulations are in current prevalent use in medical and other allied health professions including nursing (Lewis, Davies, Jenkins & Tait, 2001; Vyas et al., 2011; Vyas et al., 2010; Cook & Triola, 2009; Kane-Gill & Smithburger, 2011; Wilson et al., 2010), pharmacy (Vyas et al., 2010; Vyas et al., 2011; Cook & Triola, 2009; Sitzmann, 2011; Kane-Gill & Smithburger, 2011) and audiology (Wilson et al., 2010). They form a large basis of medical and nursing training today. The decision to use anatomical simulations (for nursing or surgery) or virtual patients (medical or audiology) depends on the profession, if it requires learning from dynamic and accurate patient responses (Wong, 2004; Dawson et al., 2004; Siemers et al., 2009; Sitzmann, 2011). This is reflected in recent simulations research in healthcare, where games, anatomical and virtual patient simulations predominate more successful training simulations in recent times and are more frequently used (Sitzmann, 2011; Issenberg et al., 2005).
Virtual patient simulations represent the ideal combination of the advantages of SP testing and computer-based simulations. High-fidelity virtual patient software can provide all the advantages of using computer-based simulations, with added strengths pertaining to SP testing in regards to realism, engagement and personal interaction. Virtual patient simulations are problem centered, and can engage the users to utilise their prior knowledge, whilst allowing for new knowledge to be created (Botezatu, 2010a; 2010b). Therefore, virtual patient software may well represent the mode of choice for use in audiology. Within audiology, virtual patient simulators have been introduced for training student audiologists, particularly with regard to pure-tone audiometry and masking. Some university programmes in the United States, Australia, Switzerland and England have started using simulators and integrating them into the curriculum (Wilson et al., 2010; Ashman, 2003; Kompis, Steffen, Caversaccio, Brugger & Oesch, 2012). Early evidence show that students take well to these simulators.

**Theoretical models of virtual patient simulator use**

There is limited outcome data on the reproducible effectiveness of simulators, and this extends to virtual patient simulators (Botezatu, 2010a; Issenberg et al., 2005; Sitzmann, 2011; Kompis et al., 2012). Different variables have been studied and few clear results have been established. Outcomes are always difficult to quantify; whether the outcome is learning, knowledge, motivation, or confidence: each research produces different outcomes. Meta-analyses have been useful for delineating common benefits and challenges (Sitzmann, 2011; Issenberg et al., 2005). There is also limited definitive evidence in areas of interest such as retention of learnt outcomes and skill transference from training into real-life clinical situations. Another key area insufficiently explored is if the difference in the student’s year of study can moderate any learning effects of simulator use. Dawson & Gould (2007) suggest that learning
outcomes may vary depending on where the students are in their training. Novice users can develop proficiency with techniques and equipment to attain a certain level of skill prior to meeting patients; while more experienced clinicians can “refresh and maintain old skills and learn and perfect new ones, plan alternative methods of interventions and trial new devices” (Dawson & Gould, 2007, p.1672). This can influence the appropriate starting point to introduce simulator use into the curriculum. There is constant call for greater research into virtual patient simulator use for learning purposes and the production of more realistic simulations to enhance these learning outcomes (Wilson et al., 2010). Despite this, previous research still provides a theoretical basis of exploring learning outcomes via virtual patient simulators.

Learning occurs internally and therefore cannot be directly measured (Anderson & Warren, 2011). Hence, indirect measures of learning such as performance curves and written assessments are used as indicators in assessing learning (Anderson & Warren, 2011). Learning theories closely related to simulator research have included the use of knowledge-based systems (KBS) (Smedley & Sutton, 2007), Affective-Behaviour-Cognitive (ABC) theory (Sitzmann, 2011), and its closely related Adaptive Characterisation of Thought theories of ACT* and ACT-R (Smedley & Sutton, 2007).

Knowledge based systems facilitate the capture and reusability of expert knowledge (Smedley & Sutton, 2007) and have been shown to be relatively effective in transferring information (Gregor & Benbasat, 1999). To transfer expertise to less-knowledgeable decision makers requires the use of explanation facilities within KBS to enhance user understanding of how the expertise in the system works (Arnold, Clark, Collier, Leech & Sutton, 2004; Smedley & Sutton, 2007). However, this model of knowledge transfer from an expert to a novice is best
suited to mentorship or apprenticeship. As such, this is not specific to audiology students intending to develop clinical reasoning skills (Cook & Triola, 2009).

An Affective-Behavioural-Cognitive (ABC) model of learning suggests that learning may take place through a combination of several aspects within a learner’s psyche. Affective components (e.g. Motivation, confidence, self-efficacy and effort) can influence the decision to participate and the quality of participation in training methods. Affective elements explored by Wilson et al. (2010) and Deladisma et al. (2008) showed consistent improvements when coupled with simulator use. Behavioural components are often goals; and successful training transfer represents the desired outcome of training. This can include areas of procedural knowledge and training transfer (Sitzmann, 2011). Finally, cognitive measures (e.g. Declarative and retentive knowledge) are important and need to be included since it forms the most common measurement for assessing learning. Declarative and retentive knowledge can usually be quantified and assessed, and are hence frequently used learning indicators.

Anderson and Warren (2011) applied the ABC model for simulation-based training in neonatology. They contend that simulations grounded in sound educational theory and require deliberate practice can be useful for both novice and experts by simultaneously improving and increasing the retention of clinical skills (which they define as comprising of cognitive, behavioural and technical skills). Anderson and Warren (2011) also suggest that this is best supported by also requiring users to reflect and attend facilitated debriefing sessions post-simulations. Their application of Bloom’s taxonomy of the cognitive domain for simulation learning is based on knowledge, comprehension, application, analysis, synthesis and evaluation (Anderson & Warren, 2011, p.60). Similarly Vyas et al. (2011) used Bloom’s taxonomy to teach pharmacy students. Vyas et al. (2011) are more realistic about the teaching process, and insist
that it is unlikely that any easy to use, high fidelity simulator is still able to substantially achieve a complete learning process as a standalone unit. This realistic consideration should be applied to fields such as Audiology, where it is unlikely that any personal clinical interactions can be solely learnt through a computer simulation.

Meta-analytical research by Sitzmann (2011) with simulation games provides the greatest recent support of ABC theory for simulation training. Sitzmann found that procedural, declarative and retained knowledge increased in a range of instructional settings through the use of computer-based simulations. Training transfer was also found to have significantly improved when compared to control groups (Sitzmann, 2011). The strongest differences were found when there was unlimited access to the simulator, the training material was actively incorporated into the simulator, and also when other training material was used to supplement the simulator. It is important to note that actively engaging training material was shown to demonstrate a more pronounced learning effect than simulation games when it was the stand-alone form of instruction (Sitzmann, 2011). However, this is not always possible with New Zealand audiological training. In an ideal setting, each student will receive actively engaging and specialised instruction as required from clinical educators and external placement sites. However, the difficult in obtaining such ideal placements means that the next most effective tools, including well-designed virtual patient simulators, are seriously considered and utilised.

Another close iteration of the ABC model is the ACT* theory. Herz and Schultz (1999) utilised this theory to investigate procedural and declarative knowledge components in accounting students within a computerised simulation setting. They found that for accounting, with its high dependence on procedural knowledge, the computer simulation was able to enhance procedural knowledge. This supports the use of ACT* theory within the accounting field,
however, it may not be generalised into allied health. Similarly, Adaptive Characterisation of Thought-Rational (ACT-R) theory (Smedley & Sutton, 2007; Anderson, 1993) is a theory of cognitive skill acquisition which involves three memory units (working, procedural and declarative memories). It is worth exploring if the three cognitive memory units are retained in practice.

**Learning outcomes of virtual patient simulator use**

Within these theoretical models, there have been several learning outcomes used to quantify benefits of virtual patient use. Recent research has focused on affective elements such as student opinions and motivations as outcome measures in virtual patient cases (e.g. Wilson *et al.*, 2010; Deladisma *et al.*, 2008). Others have explored areas of behavioural and cognitive development applied to simulator research (e.g. Sitzmann, 2011; Botezatu, 2010a; 2010b; Cook & McDonald, 2008; Kerfoot & Brotschi, 2009; Ashby & Crossley, 2010; Smedley & Sutton, 2007). Still others have focused on users’ feedback and analysed their use and desired learning outcomes to provide ideal virtual patient software design models (Huwendiek *et al.*, 2009; Issenberg *et al.*, 2005; Sitzmann, 2011). Despite this, there appears to be limited reported validity overall, requiring clearly defined and achievable learning outcomes pertaining to virtual patient use. The combination of validating learning models, providing effective feedback and refining simulator designs is crucial to advance research in this area (Willaert *et al.*, 2012).

Recent research has shown virtual patient simulators to be largely effective at achieving affective learning outcomes (Wilson *et al.*, 2010; Deladisma *et al.*, 2008; Sitzmann, 2011). Within audiology, the most recently validated work is Wilson *et al.*’s (2010) exploration of students’ opinions and motivation. After learning pure-tone audiometry, speech audiometry and acoustic immittance testing through the use of Standardised Patients (SPs) and Parrot’s Software
virtual patient simulator (Audiology Clinic with Generator v2.03), twenty-five student audiologists’ impressions of learning via both methods were investigated via questionnaires. Students felt that both methods improved their ability to interact with clients and perform basic audiometry due to the relative realism of both learning tools, the opportunities to practice skills without fear of making mistakes or harming clients, as well as by providing more hands-on experience by starting and completing each case. Students were less anxious with virtual patients than with SPs, which is unsurprising since there is less cliental risk. However, these students preferred the SP overall, citing that although the virtual patient provided good avenues to practice rare cases, the software was less realistic in its representation of real cases. Similarly, Deladisma et al. (2008) used a virtual patient simulator to teach 23 medical students case history taking and communication skills. These students also rated the use of the simulators as a positive and worthwhile experience. Once again, the determinant factor of whether students recommended future use of the simulator hinged on their perception of the virtual patient’s realism. Sitzmann’s (2011) meta-analyses demonstrated that when assessed as a whole, self-efficacy could be shown to improve through simulator use when accompanied by good feedback. This was consistent with the findings of the previous two studies. Based on these investigations, virtual patient cases would need to have greater fidelity and realism, and provide better feedback mechanisms at the end of each case (Deladisma et al., 2008; Wilson et al., 2010; Sitzmann, 2011).

Behavioural elements have not been explored in great detail through the use of virtual patient software. Early simulator research investigated the procedural knowledge improvements

---

4 It is worth noting that Audiology Clinic utilises two-dimensional virtual patients in their simulations, and speech audiometry testing in this case refers to speech reception threshold testing (SRT), which is not a standard speech audiometry test procedure in New Zealand.
associated with simulator use, which has been understudied in recent years (Schneider, Rittle-Johnson & Star, 2011). Procedural knowledge refers to information about how to perform a task or action, and awareness of the operators and conditions under which they can be applied to reach certain goals (Sitzmann, 2011, p.496; Anderson, 1993; Schneider & Stern, 2010).

Procedural knowledge can be assessed through procedural activities, whereby abstraction process inherent to the test process can be used to determine the level of procedural knowledge (Smedley & Sutton, 2007). These abstraction skills can be built into simulator designs which encourage users to draw conclusions and refer to other material to enhance their procedural knowledge (Smedley & Sutton, 2007). Flexible measurement procedures exist for procedural knowledge. For their study on senior accounting students, Smedley & Sutton (2007) assessed procedural knowledge through multiple-choice scales administered via a computer.

Procedural knowledge about how to perform tasks should be complemented with declarative knowledge as critical components of skill proficiency (Schneider et al., 2011). Schneider et al. (2011) argues that procedural knowledge should not be a simple representation of the steps taken to complete a task, but rather should represent the inherent method of arriving at the same goal using a well-reasoned algorithm of steps. Deep procedural knowledge, according to Schneider et al. (2011), should be flexible, and students must have in-depth knowledge of procedures for true skill proficiency.

Another area understudied for virtual patient research is training transference of learnt skills. Training transfer refers to the successful application of the skills gained in a training context to the job (Baldwin & Ford, 1998; Sitzmann, 2011, p.496). There is early evidence that skills gained from virtual patient use can be transferred onto real patients in a clinical setting (Botezatu, 2010; Cook & McDonald, 2008). Training transfer is an important measure of any
skill-based training. Especially for professionals, it is important to be able to apply learnt knowledge into a practical setting. Audiologists must not only be proficient with theoretical concepts; they should be able to apply these concepts when required. It would be useful to explore if this is also evident through the use of a speech audiometry simulator.

Cognitive elements have been explored across numerous studies, but there is limited validated outcome data available within virtual patient simulator research. Behavioural elements are useful to develop skill proficiencies, but more in depth knowledge is required for deeper understanding. Although speech audiometry is primarily a clinical exercise, it is important for clinicians to understand the reasons and purposes for its use. This is applied when explaining results to clients, when reporting to other stakeholders, and for justifying its use as part of the clinical test battery. Therefore, a procedural knowledge increase must be coupled with a declarative knowledge increase. There is evidence that procedural and declarative knowledge areas may interact during simulator training (Ashby & Crossley, 2010; Herz & Schultz, 1999; Smedley & Sutton, 2007). However, some research has found that when learning tasks involve both intense procedural and declarative components, one knowledge area may sometimes increase at the expense on the other (Ashby & Crossley, 2010; Smedley & Sutton, 2007). It is unknown if this relationship extends to over an extended period post-simulator use.

Declarative knowledge refers to trainees’ memory of the facts and principles taught in training and the relationship among knowledge elements, which can comprise general and abstract conceptual knowledge, as well as knowledge about specific and concrete instances and events (Kraiger, Ford & Salas, 1993; Sitzmann, 2011; Schneider & Stern, 2010). Declarative knowledge is usually the learning outcome associated with passive learning material.
Examinations of knowledge are usually based on declarative principles, and these are still frequently used within the Master of Audiology programme.

The previous areas of knowledge are all important, but they remain of no use if they cannot be recalled when needed later. Retentive knowledge is a delayed assessment of declarative knowledge and refers to trainees’ memory of the factual information taught in training several weeks or months after leaving the training environment (Sitzmann, 2011, p.496). Within a clinical setting, retained knowledge can be described as a critical precondition for knowledge to result in a substantial improvement in clinicians’ behavior (Botezatu, 2010b; Kerfoot & Brotschi, 2009). Knowledge gained by learning needs to be retained and recalled when prompted, if not knowledge is transient, and does not provide long-term benefit to clinicians, and consequently, for patients.

Simulator use has been found to directly affect retained knowledge (Sitzmann, 2011). When coupled with multiple learning strategies and good simulator design (Sitzmann, 2011; Issenberg et al., 2005), knowledge retention for cognitive areas is expected. Hence, retained knowledge has been previously narrowly defined as the retention of declarative knowledge (Sitzmann, 2011). However, using a broader theoretical framework, it would be more beneficial to explore the retention of all three previously identified knowledge areas (training transfer, declarative and procedural knowledge). If change can be observed with behavioural measures such as training transfer and procedural knowledge, then this would represent an added theoretical and practical strength of virtual patient simulator use.
Best Practice design principles for Virtual Patient Simulators

Areas of medical research and clinical education have increased their use of human simulations and virtual patient simulators which has led to a discussion of the Best Practice research principles necessary for good virtual patient simulation design (Huwendiek et al., 2009; Issenberg et al., 2005; Cook & Triola, 2009; Botezatu, 2010a; 2010b; Sitzmann, 2011). In particular, Issenberg et al. (2005) provided virtual patient simulator advice for the Best Evidence Medical Education (BEME) collaboration. In their investigations, they intended to establish the uses and features of high-fidelity medical simulations that led to effective learning. The overall outcome in professional fields should be a professional proficiency in the requisite skill area accompanied by more rigorous clinical reasoning skills (Cook & Triola, 2009). Common themes of effective virtual patient simulator design can be summarised below:

1. Feedback needs to be the most important feature of virtual patient software, and has to be provided at the end of each case. Learning should include feedback systems and components. Particularly with simulator design, feedback models should be included in the system inherently so that students are able to learn from their mistakes (Issenberg et al., 2005; Huwendiek et al., 2009; Wilson et al., 2010; Shute, 2007; Anderson & Warren, 2011).

2. It allows for repetitive practice with good access (Issenberg et al., 2005; Wilson et al., 2010; Sitzmann, 2011).

3. The virtual patient software should be integrated with the curriculum (Issenberg et al., 2005; Sitzmann, 2011; Cook & Triola, 2009).
4. Multiple learning strategies need to be used for simulator learning, and the simulator should not be the sole method of instruction. Virtual patient use can be spread out over time and does not have to be introduced purely as a block-course in a curriculum (Issenberg et al., 2005; Wilson et al., 2010; Sitzmann, 2011; Botezatu et al., 2010b; Cook & Triola, 2009).

5. Virtual patient simulator should capture a wide range of clinical variation and rare cases (Issenberg et al., 2005; Huwendiek et al., 2009; Sitzmann, 2011; Cook & Triola, 2009).

6. Virtual patient software should provide a controlled environment where users can make, detect and correct mistakes without adverse consequences (Issenberg et al., 2005; Huwendiek et al., 2009).

7. Individualised learning where participants are actively engaged in the learning process and not passive participants, with reproducible standardised educational experiences (Issenberg et al., 2005; Huwendiek et al., 2009; Sitzmann, 2011; Cook & Triola, 2009).

8. Authentic and clearly stated goals achievable through designed outcomes, with frequent assessments (Issenberg et al., 2005; Huwendiek et al., 2009; Wilson et al., 2010; Botezatu et al., 2010b; Cook & Triola, 2009).

9. Differing difficulty levels should be utilised across different cases, at appropriate difficulty levels (Issenberg et al., 2005; Huwendiek et al., 2009; Sitzmann, 2011; Cook & Triola, 2009). To promote a more flexible thinking process, simulators should be able to
improve the user’s ability to cope with novel situations (Anderson & Warren, 2011; Cook & Triola, 2009).

10. Validation of the simulator, including reviewing all aspects relating to each case
Issenberg et al., 2005; Huwendiek et al., 2009; Sitzmann, 2011; Botezatu et al., 2010b; Cook & Triola, 2009).

Having discussed ideal simulator designs, is it necessary that virtual patient software includes as much multimedia as possible and includes every conceivable element in order to create a realistic and effective simulation? Cook and Triola (2009) argue against this. A virtual patient may be distracting if there is some audio or some visual discrepancy, which can increase cognitive overload for the user, particularly during the early or learning stages of the simulator. Thus, an engaging virtual patient software design which is not over-stimulating may prove to be the most effective solution (Cook & Triola, 2009; Issenberg et al., 2005). Getting the design correct can greatly influence all measured outcomes.

Finally, although it is recognised that empathy and other related personality traits are crucial for positive client-patient experiences, and that these cannot be incorporated into the use of simulators, the positive learning aspects will only aid the able clinician to be more proficient through high competence of skills (Wilson et al., 2010; Sitzmann, 2011).
Summary and research questions

The aim of the research is to validate the speech audiometry simulator developed at the University of Canterbury against learning outcomes important to clinical audiology previously identified. Utilising these in the virtual patient simulator’s design, it is expected that an engaging learning simulator can be developed according to best practice principles.

The simulator validation needs to be according to the previously described frameworks for virtual patient simulation for healthcare training (Cook & Triola, 2009; Sitzmann, 2011). Most of the theories presented use an iteration of the Affective-Behavioural-Cognitive (ABC) model as a basis of identifying learning outcomes. There is evidence that the use of this model to validate learning outcome of virtual patients simulators is well supported (e.g. Botezatu, 2010a; 2010b; Cook & Triola, 2009; Sitzmann, 2011).

To advance clinical reasoning skills, behavioural and cognitive areas of procedural and declarative knowledge need to be well formed. This knowledge needs to be retained after training, and be applied in a real-life clinical setting. Hence, a key aspect of this study will be assessing a combination of behavioural and cognitive learning outcomes. Hence, two behavioural (procedural knowledge and training transfer) and two cognitive (declarative and retention knowledge) outcomes were assessed between simulator users and a control group.

To further advance knowledge in this field, advanced and novice students of audiology will be recruited in two studies. If different learning outcomes are observed in each group, this will also provide a more realistic indication of when to introduce a virtual patient simulator for audiological learning.
If core areas of knowledge (procedural, declarative and retentive) could be increased through the use of the Speech Audiometry virtual patient software, and if these skills can be transferred from training to real clinical situations, then this would indicate strong validity of the simulator and encourages its specific use and implementation for New Zealand audiological training.

**Hypotheses**

For the two studies, that is with advanced (AAS) and novice audiology students (NAS), the following hypotheses can be reasonably identified from the reviewed literature. It is hoped that statistically significant effects of moderate size can be obtained for each of these four hypotheses within and between groups:

1) **Training transfer is expected to increase more for AAS simulator users than AAS non-simulator users.**

   All AAS are expected to show improvements over time in their ability to transfer learnt skills from their coursework into real clinical situations. If the simulator is expected to provide a benefit towards learning clinical skills, then it is expected that simulator users should have higher rates of transference of learnt skills to a real-life clinical setting. NAS are not included in this measure since they have no prior clinical experience with speech audiometry, and cannot be validly assessed using a baseline assessment within a clinical situation.

2) **Declarative knowledge is expected to increase more for simulator users than non-simulator users.**
The ability of students to verbalise pertinent facts and or processes (Herz and Schultz, 1999; Sitzmann, 2011) should increase for all users regardless of intervention. If simulator users experience benefit, than it is expected that simulator users’ declarative knowledge should be higher than non-simulator users for both AAS and NAS (Cook & Triola, 2009; Botezatu, 2010b).

3) **Procedural knowledge is expected to increase more for simulator users than non-simulator users.**

For AAS, procedural knowledge is expected to increase if simulator benefit is found. However, given the possible inverse relationship between procedural and declarative knowledge (Ashby & Crossley, 2010; Herz and Schultz, 1999), it is possible that a decrease in procedural knowledge may be found if declarative knowledge increases. This may be more evident for novice users. Despite this, NAS simulator users are also expected to increase by more non-simulator users even if declarative knowledge gains are found.

4) **Retention of training transfer, declarative and procedural knowledge is expected to be higher for simulator users at 4 weeks post-intervention.**

Use of the simulation is expected to reinforce and enhance learnt knowledge (Sitzmann, 2011; Botezatu, 2010b). It is expected that simulator users are more able to retain learning outcomes than non-simulator users four weeks post-intervention. A counter-argument is that a high reliance on simulator use in an intense period may in fact lead to poor knowledge retention once access to the simulator is removed. It is worth exploring if this is observed in both studies of AAS and NAS.
Development of simulator and instruments

Speech Audiology Simulator Development

Design principles identified by the literature review were integrated into the consultation process. The software was written in C++/C# by a team of developers at the HIT Lab in conjunction with the Department of Communication Disorders at the University of Canterbury. The Speech Audiometry component of the virtual patient simulator for research purposes was designed to be a stand-alone component not requiring the user to complete pure-tone audiometry or case history testing. This enabled more accurate verification and validation of the speech audiometry component for this research. Written instructions were provided to simulator users explaining how to use the simulators, along with contact details in case there were difficulties. Users required a password access to use the simulator, which were provided as per the study’s requirements.

Virtual patients were available as male and female avatars representative of 6 cultural groups. 25 cases were written for the simulator, incorporating previously validated elements. The simulator was designed to perform 7 out of the 9 tasks of speech audiometry identified in the literature review. Two components, giving the client instructions and providing feedback, were not built into the simulator because it was deemed unnecessary, and to ensure that each case could be efficiently completed. The version of the simulator used for this study was limited to 15 cases of varying difficulties. Each case consisted of a short summary of each patient’s history and complaints, otoscopic results and their completed audiogram. Based on this information, users had to perform the speech audiometry testing as they would in a clinic setting. Stimuli, presentation levels and transducers had to be selected based on given information. The dashboard
contained toggles controlling the use of a CD player with a play/pause, track rewinding and forwarding buttons. A screenshot of the simulator during testing is provided below.

*Figure 2: A screenshot of the Speech Audiometry Virtual Patient Simulator during testing*

After selecting the desired word list, the corresponding track would start and the client would make appropriate responses based on their expected score for a given ear, intensity and masking level. This information was modeled into a performance-intensity function programmed for each ear, which users had to obtain through the speech audiometry procedure. The software was robust enough that a random combination of words could be used to comprise the desired score at any point, to create a realistic response set for each patient.
The simulator played and displayed the CD recording of the AB word list; however, at the time of the study, recorded voices were unable to be installed as virtual patient responses. Thus, patient responses were displayed on screen, with a variable delay after each word to mimic real patients, to encourage the user to pause as necessary. Scoring the patient’s response was performed on provided AB wordlists scoring sheets. The totaled response had to be entered into the simulator and plotted on the on-screen speech audiogram. When the user had completed their testing, they could nominate the half-peak level for each ear and its consistency with the patient’s audiogram. They could then submit their results for evaluation and were presented with the correct result. Since a test-retest reliability error of ± 3% was programmed into each patient for realism, users were told that if their answers were within 5% of the correct answer, they had correctly completed the case.

Simulator software was installed on Windows-based desktop PCs minimally running Windows XP or later, with an Intel Core 2 Duo processor with at least 2 GB or ram. The recommended monitor screen sizes are between 17 and 21.5 inches. The software was installed on 5 computers in one computer lab to ensure full, flexible and continuous access for simulator users.
**Development of assessment measures of learning outcomes:**

Since there are no universally used declarative, procedural, training transfer and retention measures currently in audiology, other measures were adapted for this study and pilot tested.

**Training Transfer**

Previous training transfer research have used itemised scoring procedures in a real-life setting as a measurement basis for training transfer, and this study proposes to offer more control and consistency in this area by proposing the use of Standard Patients (SPs) for assessment. This method of assessing students on the same tasks in a simulation setting as required in a clinical situation is well established (Bell & Kozlowski, 2008; Botezatu, 2010a; Kane-Gill & Smithburger, 2011).

Training transfer could not be evaluated by simply modifying a scale used in previous research. Speech audiometry, being a specific form of testing within Audiology, requires its own unique measure for evaluation. NZAS has developed an assessment procedure for the AB-CVC wordlists which is adapted by the University of Canterbury and the University of Auckland. Assessments are usually made through the use of Observed Structured Clinical Examinations (OSCEs), which provide a clear marking criteria for the assessor measured against one SP, which reduce the influence of tester and patient bias. A second examiner can also be used to compare inter-rater agreement, which culminates in a fair and consistent approach for testing students.

The OSCE marking schedule developed for assessing speech audiometry at the University of Canterbury was modified to assess training transfer for Master of Audiology students. To better account for simulator use, and to more accurately incorporate the research question, the marking criteria was modified. These modifications ensured a more detailed
scoring procedure, which allowed for more subtle observations to be included (particularly with masking). Content validity was achieved through the use of experts from the Department of Communication Disorders at the University of Canterbury. Components unrelated to learning using the simulator and provided materials were excluded. The final measure was a flexible 59-item marking schedule with a total of 119 points.

Since training transfer involves previous experience of speech audiometry, it was practical only to assess training transfer with the Master of Audiology students. Hence, construct validity was provided in a predictive manner, since there is no direct method of evaluating training transfer in a more practical way. Criterion validity was established through its concurrent evaluation with cognitive components such as declarative and procedural knowledge, which prevented a single behavioural measure being the sole outcome measure of simulator use.

Participants were assessed on one SP per time period, prior to completing any of the other measures. This was done to ensure that any theoretical knowledge was not presented immediately prior to training transfer assessment testing. Three standardised participants (SP), one for each time period, were recruited. Each case was designed in such a way that masking was always required for both ears based on appropriate initial presentation levels, the scoring schedule would be robust, and the same number of items were evaluated for each SP across the three time periods. The first (baseline) SP was a male with a unilateral conductive loss (simulated with a ear plug in his left ear) with consistent results expected in both ears. The second (post-intervention) SP was also a male with a unilateral conductive loss (simulated with a ear plug in his right ear) with consistent results expected in both ears. The third (retention) SP was a female with a unilateral mixed loss, with a false audiogram for her right ear. Thus, inconsistent results were intended for both her ears.
The test procedure replicates the OSCE test procedure used as part of the University of Canterbury Master of Audiology programme. Participants were provided with a short case history and the SP’s audiogram, and given as much time as they needed before testing. Scoring commenced when the participant began to give instructions, and ended when the participant informed the SP of the end of testing. No feedback was given to participants during testing.

Reliability was to be ascertained in several ways. The main assessor (primary researcher) was used in all three time periods to ensure consistency of scoring. The assessor was expected to time the participant and score the participant using the developed criteria and intervene only in cases of equipment malfunction. To evaluate the prominence of examiner bias, a random sample of 33% of participants were co-evaluated by a second researcher at each time period. Inter-rater agreement was obtained to ensure consistency of scoring. The Cronbach alpha coefficients for internal reliability and inter-rater agreement values are reported in the results.

**Procedural Knowledge**

Procedural knowledge can be assessed using a variety of methods including the use of OSCE style testing or written tests incorporating multiple-choice or short-answer questions (Smedley & Sutton, 2007; Sitzmann, 2011). To prevent a heavy weighting of interactions with the SP, reduce multicollinearity with the training transfer instrument, and select an assessment that can suitably incorporate novice students (who cannot be assessed via a speech audiometry OSCE), the decision was made to use multiple-choice questions.

Following Schneider et al.’s (2011) contention that true procedural knowledge must not only contain the stepwise process required of the skill, but also include deep knowledge required for genuine understanding of the skill, it was necessary to improve the scope of the instrument.
Cases were arranged in order with the procedural knowledge for each task probed using a challenging multiple-choice scale. Not only did this make the scale more robust, it also ensured that a consistent scoring system remained challenging across all cases.

Using this design, procedural knowledge was assessed through the use of six cases, with two cases per time period. This increased variability in scoring and prevented early plateau effects. This also ensured that questions were probed several times to reduce the influence of chance. All participants completed the same cases at each time period.

Sixteen multiple-choice questions were asked per case across the procedural aspects and tasks of speech audiometry. Questions probed simple and deeper procedural knowledge in otoscopy, interaural attenuation, reading pure-tone audiograms, deciding presentation levels, obtaining both PI max and half-peak levels through speech audiometry as well as reading them off graphs, and ascertaining consistency of speech audiometry in relation to pure-tone audiograms. Tests were provided on paper and in colour, so clear inferences can be made from otoscopy images and audiograms. Procedural knowledge testing was always completed prior and independent to declarative knowledge testing at all times. This prevented participants from referring to another test to correct their answers.

Content validity was also achieved through the review of the questionnaire by similar experts. As with training transfer testing, one assessor marked all participants’ responses against a fixed marking schedule to ensure consistency. For reliability purposes, a second assessor verified the scoring of each participant by checking each test. Internal reliability estimates could only be provided after testing, and are reported in the results section.
**Declarative Knowledge**

Declarative knowledge can be assessed via oral examinations (Randell, Hall, Bizo & Remington, 2007), multiple-choice questionnaires (Ifenthaler, 2010; Bell & Kozlowski, 2008) or open-ended short answer questions (Schneider & Stern, 2010). For this study, a short-answer written test remains the most suitable method for assessing declarative knowledge since it is able to obtain a breadth of responses about each participant’s understanding of speech audiometry. As a different instrument, an open-ended questionnaire will allow for greater examination variety for this study. This can account for any difficulties participants experience with particular methods, and to maintain the participant’s interest during testing.

Declarative knowledge was evaluated using 12 open-ended short answer questions worth 36 marks. These included many of the nine tasks of speech audiometry previously categorized. This also allowed for evaluation of phonetic scoring capabilities of each participant in an objective manner. Declarative knowledge of speech audiometry was probed multiple times using subtle and parallel questioning methods. One declarative knowledge questionnaire was used for each time period. Three questions were repeated across the three time periods since they were core knowledge questions pertaining to the purpose of speech audiometry, presentation and masking formulae. Learning effects were more likely to occur with these questions, but this was not deemed to be a disadvantage. Questions were limited since they were simultaneously probed in other assessments (i.e., procedural knowledge and training transfer). Declarative knowledge assessments were completed last in all testing, and were independent to other tests. This was especially crucial due to the open-ended nature of the assessment, and no outside influence was desired. All participants received the same test at each time period. Content validity was ascertained in a similar way to procedural knowledge. Criterion validity remains predictive since
it is a new questionnaire specific to speech audiometry using the AB-CVC words list. As before, participants’ responses were marked by a main assessor and verified by a second assessor.

**Retained Knowledge**

Knowledge retention does not require its own scale and was measured using constructs established from the three previous criteria. Instead of comparing the final result with the baseline measure, it was more meaningful to show the change (if any) during the period immediately post-intervention compared to four weeks later. Therefore, retained knowledge will be represented by the non-significant change between the time periods from immediately post-intervention compared to 4 weeks later (i.e. the comparison between Times 3 and 2). This was consistent with Sitzmann’s (2011) meta-analysis found that retention assessments typically ranged from 2 to 6 weeks post-intervention. Therefore, a third version of each instrument was necessary as the assessment measure of retained knowledge.

Based on the research presented, the selection of this method for assessment was valid in its construct and in content (Lo, Devine, Evans, Byars, Lamm, Lee, Lowe & Walker 2011; Sitzmann, 2011). Any reliability concerns are limited to each scale’s internal reliability and maturation effects observed in participants. However, since all students undertook essentially the same study design, there is no expected difference between groups due to maturation effects.

**Validity of measures**

A common criticism of computer simulator research is the lack of validity establishment and reliability estimates provided (Issenberg et al., 2005; Cook & Triola, 2009; Sitzmann, 2011). Each instrument’s accuracy is limited by its validity in measuring what it intends to do (Bray et al., 2011). Construct validity has been demonstrated for each measure in the following
subsections by incorporating key elements of recent research and applying them into an audiological context. Clinical experts at the University of Canterbury reviewed content validity based on the criteria described in the tool (Bray et al., 2011; Kardong-Edgren, Adamson & Fitzgerald, 2010). They ensured that all relevant sections were included and probed in multiple ways; excessive items, and elements which could not be trained through the use of the simulator, were excluded. Predictive validity of each measure can only be ascertained after the study.

**Reliability estimates**

Reliability is important since it ensures that the instruments yield similar results for the targeted measures. This extends to the verification of inter-rater reliability when several participants are having their skills assessed subjectively by one or many assessors (Bray et al., 2011; Kardong-Edgren et al., 2010). Cronbach alpha coefficients and inter-rater reliability of the study are reported in the results.

**Pilot testing**

Pilot testing of the measures was performed after the development of measures through consultation with three practicing audiologists. The measures showed good practical reliability and were further refined. The audiologists’ input was sought and any confusion with question wording, testing length and marking schedule discrepancies were resolved. These testers also practiced on the simulator and noted inconsistencies which needed resolution and debugging. It was only after pilot testing that recruitment for the study began.
Method

Study 1: Advanced Audiology Students (AAS)

Participants

19 of the 22 (86%) Master of Audiology (MAud) students enrolled at the University of Canterbury participated. Students had previously been taught theoretical and clinical speech audiometry, with senior students expected to be more proficient. A simple incentive was provided, and the students neither gained course credit, nor were penalized for withdrawing or refusing to participate in the study. All students had previous completed 2 hours on the pure-tone audiometry section of the virtual patient simulator for their coursework.

Materials

All participants

Advanced Audiology Students (AAS) were provided with a copy of the Speech Audiometry protocols obtained from the 2012 University of Canterbury Speech & Hearing Clinic Protocols. This 4-page section details the test procedure and interpretation of speech audiometry results speech audiometry. Notes of the presentation given to Novice Audiology students (NAS) was also provided to all participants. AAS participants were not required to attend the lecture as their current knowledge and coursework had exceeded the lecture content. All material was provided to each participant in print and electronic form 24-hours prior to each participant’s first session. An information sheet for the study containing their rights and responsibilities, contact information of the main researcher, and the AB-CVC words lists were also provided for their reference.
**Simulator users**

Participants who were later selected as simulator users were subsequently provided with a one-page instruction manual on how to use the simulator, detailing their username and passwords. This username and password was only usable during the period of stipulated simulator use for the user (i.e. the two-week period of intervention assigned to the simulator user). Simulator users were also asked not to use the simulator in the presence of other participants who were in the non-simulator group.

**Equipment**

Standardised Patient (SP) testing was completed in one of two audiological booths on campus. GSI –61 audiometers still in calibration were used. Speech audiometry was administered using the Millennium edition of the AB-CVC Words list via an attached CD player. Participants had to calibrate the CD output at the start of testing as per the complete speech audiometry testing procedure. Participants could choose to present speech stimuli with either TDH-39 supra-aural headphones or ER3-A foam-tip inserts with tube earphones for each patient. A Casio stopwatch was used to time each participant from when they gave the SP instructions till the conclusion of test. The assessor(s) were seated beside the participant in a position where they could see the participant’s actions but were not interfere with the test process.

AAS simulator users could practice on four computers at the Master of Audiology students’ area on campus. 24-hours access to these computers was provided, and printed copies of the AB Words lists was also supplied beside each computer. Users had to use their secure passwords to access the simulator over the 2-week period of intervention, and complete a minimum of 5 cases. This was verified via data logging of the cases on the nominated computers.
General procedure

Study 1: Advanced Audiology students

A double-repeated measures intervention design over six weeks was utilised. A baseline measure was followed by a post-intervention assessment after two weeks, and was subsequently followed by a retention assessment following a period of non-intervention (four weeks). This is summarised in Table 1 below.

Table 1: Research design and time frames

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline Assessment (Time 1)</th>
<th>Intervention Period (2 weeks)</th>
<th>Post-intervention assessment (Time 2)</th>
<th>Non-instruction period (4 weeks)</th>
<th>Retention Assessment (Time 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Advanced Students) Simulator</td>
<td>a</td>
<td>s</td>
<td>a</td>
<td>x</td>
<td>a</td>
</tr>
<tr>
<td>Non-simulator</td>
<td>a</td>
<td>n</td>
<td>a</td>
<td>x</td>
<td>a</td>
</tr>
<tr>
<td>Study 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Novice students) Simulator</td>
<td>i, b</td>
<td>s</td>
<td>b</td>
<td>x</td>
<td>b</td>
</tr>
<tr>
<td>Non-simulator</td>
<td>i, b</td>
<td>n</td>
<td>b</td>
<td>x</td>
<td>b</td>
</tr>
</tbody>
</table>

Notes:

a  Participants assessed using standardised patient for training transfer, and written tests for procedural and declarative knowledge
b  Participants assessed using written tests for procedural and declarative knowledge only
i  45-minute introductory lecture to speech audiometry for participants prior to testing
n  Participants were asked to read and learn through the presentation notes and the UC Protocols
s  Participants were also asked to read and learn through the presentation notes and the UC Protocols, but were also given instructions to use the simulator and complete a minimum of 5 virtual patient cases on the simulator.
x  All participants were not allowed to use the simulator, and not asked or required to read any provided material
Prior to the baseline measure, participants were provided with the clinical protocols and lecture notes. Informed consent was obtained prior to the first session, and responsibilities pertaining to the study were explained and clarified.

For the baseline measure, each participant was individually assessed over two days. Participants were asked not to communicate with anyone else about the study. At each session, training transfer was assessed prior to procedural knowledge and declarative knowledge. After all participants had been tested, they were divided into two groups with an equal number of first and second year Master of Audiology students in each group. Repeated measures analysis meant that baseline score matching was not as crucial for computing one-way ANOVAs. Simulator users were then given instructions on how to use the simulator and informed of that they were required to complete a minimum of five cases over the intervention period. Non-simulator users were instructed to learn through the other provided material instead.

After intervention, participants were assessed in a similar manner. Simulator users’ feedback on positive and negative aspects of the virtual patient simulator was also sought at this point. Subsequently all participants were informed that they were not required to undertake any compulsory learning, but would be re-assessed in four weeks time. Simulator users were now unable to use the simulator software.

At four weeks post intervention, all participants were re-assessed for retained knowledge. Incentives for completing the study were paid out on this session. All participants were informed that if they wished to use the simulator at the conclusion of the study, they would be provided with access as required. Statistical analysis was computed using IBM SPSS Statistics 19 for Mac using a series on ANOVAs and one-way repeated measures for each variable in each study.
Descriptive statistics, reliability and inferential statistics for the hypotheses are reported in the results section.

**Study 2: Novice audiology students (NAS)**

**Participants**

Novice students of audiology were undergraduate students not currently enrolled in the Master of Audiology programme, and who had completed the undergraduate course CMDS 242 (Introduction to Audiology). This course is a compulsory pre-requisite of the Bachelor of Speech Language Pathology (BSLP) degree, and introduces the basic tenets of audiology, including the ear and hearing, pure-tone audiometry and case history taking.

18 of the 40 students who responded participated in the study (45%). This comprised of BSLP students and prospective Master of Audiology students who had completed CMDS 242. 1 participant dropped out before the final (retention) measure without cause. Since this participant had completed intervention assessment, they were only excluded from retention analysis.

**Materials**

**Lecture for novice audiology students**

In addition to the other material provided, a 45-minute lecture was given to NAS participants to ensure a consistent starting knowledge base for the study. The lecture reviewed introductory audiology concepts covered in CMDS242, before new information about speech audiometry and its essential components were presented. Content delivery included facts, illustrations, diagrams, and a case example where attendees had to apply newly learnt
knowledge. By attending the lecture, all participants have all required information to perform well in the cognitive areas of testing.

The same experts consulted for the other instruments evaluated content validity. The main researcher presented this lecture to all participants in five sessions at the Hit Lab, and a strict timetable and script was used to deliver the lecture. None of the participants had witnessed the New Zealand adaption of speech audiometry in a clinical situation previously, and all participants reported the speech audiometry content of this lecture as new to them.

**General procedure**

The same timeframe with a slightly modified general procedure was used for NAS participants. The key differences was that a training transfer measure (i.e. A speech audiometry OSCE assessment) was not used, and prior to the baseline measure, each participant attended one of the five available lectures and was given a few minutes at the end of the lecture to briefly review the presented material. All measures comprised of the procedural knowledge test followed by the declarative knowledge test. To further reduce presenter bias, participants from each lecture session were later divided into simulator and non-simulator users. Both the simulator and non-simulator groups were provided with the same instructions as per the study of AAS participants. Five computers were made available with 24-hours access for NAS simulator users to complete their five minimum cases.

**Research Design**

This quasi-experimental design allows for the contribution of simulator use towards the forms of knowledge to be assessed. A change in score from pre- and post-intervention will allow for more accurate demarcation from one group to the next. The design does not allow for
complex controls and blinding, however, given the complexity of the design and the potentially small sample size, general findings could potentially be drawn from this study for further refinement in future research.

The intervention period was designated as two weeks. Previous research (Sitzmann, 2011; Cook & Triola, 2009; Wilson et al., 2010; Issenberg et al., 2005) suggests that the length of time for learning should be reflective of learning goals. Learning goals in the areas of procedural, declarative knowledge and training transfer can be reasonably achieved in an intensive two-week intervention period. Next, the workload requirements of the target population meant that any increase in training duration may have resulted in lower participation rates, and given the time intensiveness of the study, a high participation mortality rate was not risked (Schiavetti, Metz & Orlikoff, 2010). Mortality, in this case, refers to participants dropping out between measures (Schiavetti, Metz & Orlikoff, 2010). The retention period of four weeks was selected due to similar concerns about participation mortality. Simulator users were prevented from accessing the simulator during this period to assess the true retention of simulator-gained knowledge. Participant mortality was minimal with an overall completion rate of 97%.

There are three main disadvantages that should be acknowledged with this design. Firstly, as previously explained, participation mortality can be a concern. Secondly, learning effects or maturation can be observed between time periods due to the similar structure of testing. Thus, a plateau effect is more likely to be expected, particularly amongst audiology students. This can be somewhat controlled in several ways. Variability in content and cases between time periods for the measures meant that participants could be asked variants of the same questions. This removed learning effects based purely on memorisation. Further, the time periods between testing, the number of questions asked, and the difficulty of each measure combined to negate
possible learning effects. Strict rules of discussing this study with another participant were imposed with voluntary approval. Simulator users were not to practice in the presence of other participants, especially non-simulator users. Finally, there can be difficulty attracting participants due to the length of the study. However, the strengths of the study design overcome this potential loss of sample participants. Incentives need to be attractive enough to overcome any perceived challenge in committing to a long study. In this case, a major prize in a prize draw was provided in addition to a modest incentive to encourage participation and completion of the study.

**Funding**

Funding for the project was provided by three means. The ReSound Research Grant provided monetary support for participants’ incentives. The Tertiary Education Commission (TEC)-TSI grant for Immersive Education in Audiology provided scholarship and course fees support for the main researcher. Finally, the HIT Lab was supported through its own funding sources to produce a virtual patient simulator for audiology. This ensured a high-functioning programming team to create, develop and maintain the software package. No conflicts of interests are declared.

**Ethics**

Ethics approval was sought through the University of Canterbury’s Human Ethics Committee Low Risk process. No deception was used, and participants were free to withdraw at any time without penalty. Approval was granted (HEC 2012/20/LR) after revision of participant incentives, and no restrictions were stipulated.
Results

Descriptive Statistics

Demographic information for participants is summarized in Table 2 below. For the AAS study, both student years were evenly split between simulator and non-simulator groups. The age spread for AAS participants is a direct reflection of students in the audiology programme. The relatively few males enrolled in the audiology course as well as in BSLP programmes (under which CMDS 242 is taught) make it difficult to match participants by gender. The age range in the NAS study is comparatively younger, which is unsurprising given that these students are still completing undergraduate degrees. Second and third year BSLP students were the primary participants. The majority of fourth year students were already on externships at the time of the study. Students who had completed CMDS 242 but who were studying other degrees were divided between the simulator and non-simulator groups.

Table 2: Summary of participants’ age, gender and year of study:

| Item   | Categories | AAS Study | | NAS Study |
|--------|------------|-----------| |           |
|        |            | Simulator (n=9) | Non-Simulator (n=9) |
|        |            | Simulator (n=9) | Non-Simulator (n=9) |
| Gender | Male       | 2          | 2 | 1         | 0 |
|        | Female     | 7          | 7 | 8         | 9 |
| Year of Study | Year 1 | 4          | 4 | -         | - |
|        | Year 2     | 5          | 5 | -         | - |
|        | BSLP200    | -          | - | 2         | 3 |
|        | BSLP300    | -          | - | 4         | 3 |
|        | BSLP400    | -          | - | 0         | 1 |
|        | Other Major| -          | - | 3         | 2 |
| Age    | <20        | 0          | 0 | 1         | 2 |
|        | 20-29      | 5          | 3 | 8         | 6 |
|        | 30-39      | 1          | 5 | 0         | 1 |
|        | 40-49      | 1          | 0 | 0         | 0 |
|        | <50        | 2          | 1 | 0         | 0 |
Summary of Results

The following table (Table 3) shows the mean group scores across both studies of both advanced audiology and novice audiology students.

Table 3: Summary of participants’ mean raw scores (standard deviations in parentheses):

<table>
<thead>
<tr>
<th></th>
<th>Training Transfer</th>
<th>Declarative Knowledge</th>
<th>Procedural Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1 (AAS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulator</td>
<td>83.22</td>
<td>91.00</td>
<td>89.78</td>
</tr>
<tr>
<td></td>
<td>(15.37)</td>
<td>(9.15)</td>
<td>(12.32)</td>
</tr>
<tr>
<td>Non-Simulator</td>
<td>81.67</td>
<td>87.00</td>
<td>90.11</td>
</tr>
<tr>
<td></td>
<td>(16.64)</td>
<td>(8.15)</td>
<td>(10.31)</td>
</tr>
<tr>
<td>Study 2 (NAS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulator</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Simulator</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All cases have n=9, except for marked boxes (*) where n=8, where one participant did not complete Time 3.

This data will be displayed graphical form, demonstrating the relationship in variables over time for the different groups. This will be shown first for the AAS participants and then for the NAS students as pertaining to the hypotheses. One AAS participant did not return for assessment at time 2 and was subsequently eliminated from all data analysis. One NAS participant did not complete the retention component of the study and was excluded from analysis only at time 3.
Advanced Audiology students (AAS)

The same analysis method was employed for each study. An ANOVA within-between subjects analysis (or split-plot ANOVA [SPANOVA]) was performed to compare the main and interactive pre- and post-intervention scores between and within subjects for the dependent variables (training transfer, declarative knowledge and procedural knowledge) across two independent variables of group (simulator/non-simulator) and time (1, 2, 3) (Pallant, 2005, p.239; Tabachnick & Fidell, 2001). This provided a broad overview for each dependent variable if there were significant main or interaction effects observed over time.

Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. Univariate outlier analysis revealed some outliers with no extreme outliers. This was accepted as a true representation of the dataset since each participant is assumed to have performed their best given the amount of work they put in.

Multivariate outlier analysis using Mahalanobis distance values was below the critical value for 9 dependent variables (Pallant, 2005, p.251), indicating that no multivariate normality and no multivariate outliers present. Scatter plots revealed linearity of dependent variables. Bivariate correlations between variables were between .043 and .625, indicating moderate correlations and no multicollinearity of dependent variables. Box’s test of equality of error variances was not able to be performed, but Levene’s test of equality of error variances was not significant, therefore equal variances are assumed.

There was no statistically significant difference among simulator and non-simulator users on a linear combination of dependent variables: F(9, 8)=.974, p=.520; Wilks’ Lambda=.477;
partial eta squared = .523. When the dependent variables were considered separately, none of the variables were significant using a Bonferroni adjusted alpha level of .0056. An inspection of the mean scores indicated that whilst the simulator group performed better after intervention, the difference in scores was relatively small.

For further analysis a series of one-way repeated measures were used to explore if an intervention difference was found for each learning outcome in each group. Repeated measures analysis was done with the following assumptions: the data was interval, participants were randomly sampled, observations were mostly independent from each other, the data was distributed normally and variances were homogenous (Pallant, 2005). The Shapiro-Wilk statistic, used for small samples, was not significant for all variables when divided into simulator and non-simulator groups, indicating normality of each learning outcome. Statistics for Levene’s test of equality of variances and Mauchly’s test of sphericity are reported for each hypothesis. Knowledge retention was assessed by its own method and is reported under its respective subsection. Finally, reliability statistics were computed for each scale. Graphical plots representing the groups’ mean scores (Table 2) are displayed at the end of each learning outcome.

Training Transfer for AAS

The means for both groups’ training transfer increased post-intervention (see graph). However, at Time 3, the AAS simulator group regressed whilst the non-simulator group improved.

Mauchly’s Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = .429, p = .807$. The main effect within each group across time was statistically significant [$F(2,32)=3.745, p=.035$, partial eta squared=.190] with a large effect size (Pallant,
This showed that there was a significant change in training transfer for both groups with time. However, the main effect between groups was not significant \[ F(1,16)=.136, p=.717, \text{partial eta squared}=.008 \], indicating that this improvement may have occurred regardless of the simulator use. The interaction of time and group was not significant \[ F(2,32)=.265, p=.769, \text{partial eta squared}=.016 \]. It is not surprising therefore, that significance was not reached given the small partial eta square. This means that when group differences are accounted for, the increase in post-intervention scores is not due to group assignment.

A series of one-way repeated measure between pre- and post-intervention scores was run for each group to identify specific differences with intervention. Levene’s test of equality of error variances for each time period was not significant, therefore equal variances are assumed. There was no homogeneity of intercorrelations given by Box’s M of 13.302 which was associated with a p value of .107, which was not significant at the conservatively estimated \( p<.001 \) level (Pallant, 2005, p.241).

The simulator group had a significant change in training transfer skills between post-intervention \[ \text{Wilk’s Lambda}=.598, F(1,8)=5.367, p=.049, \text{partial eta squared}=.402 \] with a large effect size. The non-simulator group did not experience a significant change across this same time period \[ \text{Wilk’s Lambda}=.888, F(1,8)=1.011, p=.344, \text{partial eta squared}=.112 \]. Therefore, it can be seen that a significant improved in training transfers skills was only observed in the simulator group.
Declarative Knowledge for AAS

A visual inspection of the means reveals a similar pattern of data as observed for training transfer. Both groups improved after a period of intervention, but while the non-simulator group continued to increase their declarative knowledge at Time 3, the opposite effect was observed for the simulator group.

Mauchly’s Test of Sphericity indicated that the assumption of sphericity held, $\chi^2(2) = 4.276$, $p = .118$. Equal variances were assumed since Levene’s test of equality of error variances for each time period was not significant and there was no homogeneity of intercorrelations (Box’s M = 9.935, $p=.247$) since Box’s M was not significant at the conservatively estimated $p<.001$ level (Pallant, 2005, p.241).
On the basis of these assumptions, there was no significant interaction between group and time \([F(2,32)=1.881, p=.169, \text{partial eta squared}=.105]\). Therefore, any differences observed were not due to group assignment. There was also no statistically significant main effect across the two groups \([F(1,16)=.958, p=.342, \text{partial eta squared}=.056]\).

There was a significant main effect with time \([F(2,32)=6.275, p=.005, \text{partial eta squared}=.282]\) with a large effect size. Post-hoc repeated measure testing with divided groups revealed that both the simulator [Wilk’s Lambda=.667, \(F(1,8)=4.000, p=.081, \text{partial eta squared}=.333\)] and non-simulator [Wilk’s Lambda=.640, \(F(1,8)=4.492, p=.067, \text{partial eta squared}=.360\)] groups failed to obtain a statistically significant improvement in declarative knowledge.

**Figure 5: Declarative Knowledge for Advanced Audiology Students**
**Procedural Knowledge for AAS**

Procedural knowledge means for both groups reveal an unusual pattern. Both groups decreased after intervention, but recovered when re-assessed at Time 3. All assumptions held as previously. Mauchly’s Test of Sphericity again indicated that the assumption of sphericity had not been violated, \( \chi^2(2) = 2.716, p = .257 \). Levene’s test of equality of error variances for each time period was not significant, therefore equal variances are assumed. No homogeneity of intercorrelations were given by Box’s M of 7.338 (p=.444) which was not significant at the conservatively estimated p<.001 level (Pallant, 2005, p.241).

For procedural knowledge, there was no significant interaction between group and time \([F(2,32)=.631, p=.539, \text{partial eta squared}=.038]\) indicating that both groups experienced similar results across time. There was no significant main effect with time \([F(2,32)=1.869, p=.171, \text{partial eta squared}=.105]\) with a large effect size, nor across the two groups \([F(1,16)=.067, p=.799, \text{partial eta squared}=.004]\). This was not surprising given the unusual pattern of results. This was confirmed with post-hoc repeated measures testing where both the simulator \([\text{Wilk’s Lambda}=.980, F(1,8)=.165, p=.695, \text{partial eta squared}=.020]\) and non-simulator \([\text{Wilk’s Lambda}=.794, F(1,8)=2.076, p=.188, \text{partial eta squared}=.206]\) groups did not significantly change after intervention. Despite the unusual decrease post-intervention, their scores were not found to be significantly different for both groups when compared to pre-intervention.
Figure 6: Procedural Knowledge for Advanced Audiology Students

Knowledge Retention for AAS

For the purposes of this study, retention is defined as the non-significant negative change between Times 2 and 3. A series of one-way repeated measures between the simulator and non-simulator groups was run for each learning outcome by comparing Times 3 with 2.

For Training Transfer, both the Simulator group [Wilk’s Lambda=.972, F(1,8)=228, p=.646, partial eta squared=.028] and non-simulator group [Wilk’s Lambda=.950, F(1,8)=.418, p=.536, partial eta squared=.050] did not change in a statistically significant way at Time 3 compared to Time 2. Therefore, training transfer was retained for both groups 4 weeks after intervention.

For Declarative knowledge, the simulator group significantly regressed 4 weeks after intervention, indicating that they failed to retain declarative knowledge [Wilk’s Lambda=.474,
F(1, 8) = 8.892, p = .018, partial eta squared = .526]. However, declarative knowledge was retained for the non-simulator group [Wilk’s Lambda = .978, F(1, 8) = .179, p = .683, partial eta squared = .022]. This unexpected finding will be further investigated in the discussion.

For Procedural knowledge, both the simulator group [Wilk’s Lambda = .854, F(1, 8) = 1.371, p = .275, partial eta squared = .146] and non-simulator group [Wilk’s Lambda = .666, F(1, 8) = 4.020, p = .080, partial eta squared = .334] showed good retention of procedural knowledge. This was largely assisted by both groups’ poorer scores at Time 2.

In summary, the non-simulator group retained all Training Transfer, Declarative and Procedural knowledge gained. The simulator group failed to retain declarative knowledge 4-weeks post-intervention.

**Novice Audiology students (NAS)**

The same analysis procedures were used for novice students. Only two dependent variables were assessed using the ANOVA within-between subjects analysis: declarative and procedural knowledge. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. The Shapiro-Wilk statistic, used for small samples, was not significant for both variables when divided into simulator and non-simulator groups, indicating normality of the dependent variables for testing. Univariate outlier analysis revealed some outliers with no extreme outliers. This was accepted as a true representation of the dataset of all participants.

Post-hoc analysis over a series of repeated measures was performed with the same assumptions as AAS participants, with no violations noted. Levene’s test of equality of variances
and Mauchly’s test of sphericity are reported for each hypothesis. Knowledge retention was assessed by the same method as the AAS study. Reliability statistics were also computed for each scale, and graphical plots of Table 2 are displayed at the end of each learning outcome.

**Declarative Knowledge for NAS**

The graph shows steep increases in the means of both groups’ declarative knowledge between pre and post-intervention. However, the simulator group declined for Time 3, whilst the non-simulator group increased. This pattern was also observed for the AAS study.

Mauchly’s Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 2.516, p = .284$. However, Levene’s test of equality of error variances for each time period was significant for Time 1 $[F(1, 15) = 5.811, p = .029]$, therefore equal variances cannot be assumed across the two groups at Time 1. The decision was made to proceed and analyse the data using repeated measures testing. There was no homogeneity of intercorrelations as given by Box’s M of 19.818 ($p = .018$) which was not significant at the conservatively estimated $p < .001$ level (Pallant, 2005, p.241).

The interaction effects of time period and group assignment for Declarative Knowledge was not significant $[F(2,30) = 4.94, p = .615$, partial eta squared $= .032]$. It is not surprising therefore that significance was not reached given the small partial eta squared. The main effect within groups of time was statistically significant $[F(2,30) = 4.509, p = .019$, partial eta squared $= .231]$ with a large effect size (Pallant, 2005, p.245; Cohen, 1988). This showed that there was a significant change in declarative knowledge across time for each group. Although the main effect between groups was not significant across the three time periods $[F(1,15) = .165, p = .691$, partial eta squared $= .011]$, further testing via post-hoc repeated measures analysis showed that the
simulator group’s declarative knowledge improved statistically significantly with a large effect size [Wilk’s Lambda=.600, F(1,8)=5.333, p=.05, partial eta squared=.400]. This increase was not significant for the non-simulator group [Wilk’s Lambda=.910, F(1,8)=.796, p=.398, partial eta squared=.090]. Therefore, only the simulator group had a significant increase in their declarative knowledge post-intervention.

*Figure 7: Declarative knowledge for Novice Audiology Students*

![Graph showing declarative knowledge for Novice Audiology Students]

**Procedural Knowledge for NAS**

The same unexpected pattern found for procedural knowledge in the AAS study was observed. Both groups decreased after intervention, but recovered when re-assessed at Time 3.

An exploration of assumptions found no violations. Mauchly’s Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = .116$, $p = .944$. 
Levene’s test of equality of error variances for each time period was not significant, therefore equal variances are assumed. There was no homogeneity of intercorrelations as given by Box’s M of 12.988 (p=.121), which was not significant at the conservatively estimated p<.001 level (Pallant, 2005, p.241).

There was no significant interaction effect between groups and time [F(2,30)=.265, p=.769, partial eta squared=.017] for procedural knowledge. However, there was a significant main effect with time [F(2,30)=13.538, p<.001, partial eta squared=.474] with a large effect size. With repeated measures testing, the simulator group [Wilk’s Lambda=.255, F(1,8)=23.376, p=.001, partial eta squared=.745] and the non-simulator group [Wilk’s Lambda=.470, F(1,8)=9.021, p=.017, partial eta squared=.530] showed significant regression in their procedural knowledge between after intervention. Although this was possible, it was not predicted. It is equally interesting to note that the recovery at Time 3 meant that there was no resulting statistically significant difference between subjects [F(1,15)=.062, p=.807, partial eta squared=.004] across the three time periods.
A series of one-way repeated measures between the simulator and non-simulator groups was then run for each learning outcome by comparing Times 3 with 2 to check for the presence of statistically significant differences.

For Declarative knowledge, both the simulator group [Wilk’s Lambda=.714, F(1,8)=3.200, p=.111, partial eta squared=.286] and non-simulator group [Wilk’s Lambda=.983, F(1,7)=.122, p=.737, partial eta squared=.017] showed good retention of knowledge weeks after the study.

For Procedural knowledge, the simulator group showed retention of their scores from Time 2 [Wilk’s Lambda=.862, F(1,8)=1.282, p=.290, partial eta squared=.138]. Consistent with
the unexpected pattern of results, the non-simulator group in fact improved significantly in their procedural knowledge in the four week period post-intervention [Wilk’s Lambda=.391, F(1,7)=10.905, p=.013, partial eta squared=.609]. This is likely due to the group’s poor scores post-intervention.

**Reliability reports**

**Training transfer**

A random selection of 6 participants in each time period (33%) was co-assessed by a second examiner. The agreements in scoring in the 59-item marking schedule was summed for each participant, and expressed over the total number of items. Inter-rater agreement percentages for the 59-items were 86.72% (standard deviation of 6.73) in Time 1, 89.27 %(s.d.=2.05) in Time 2 and 89.55% (s.d.=4.34), which indicates excellent agreement between the markers and demonstrates the robustness of the scale. All disagreements were discussed and resolved, and only the scoring of the main assessor was used in data analysis. Internal reliability of the scale was strong, with a Cronbach alpha of .715 for the 59-item scale across the 3 time periods (n=54). When broken down, Cronbach’s alpha for the 59 items (n=18) was .791 for Time 1, .535 for Time 2, and .632 for Time 3.

**Declarative knowledge**

Reliability for declarative knowledge was measured across the three time periods. For the AAS study, the declarative knowledge measure obtained Cronbach’s alpha coefficients for the 12-items of .718 in Time 1, -.480 in Time 2, and .309 in Time 3. This was inconsistent over the three time periods, with poor reliability in Time 2. The same scale, provided to NAS participants, obtained Cronbach’s alpha coefficients of .710 in Time 1, .632 in Time 2, and .646 in Time 3.
This scale was more reliable for novice students across the three time periods. More importantly, when assessed together, the effect of using the same scales across both studies meant that Cronbach’s alpha coefficients was .711 in Time 1, .539 in Time 2, and .533 in Time 3. This means that the scale was internally reliable when assessed as a whole.

**Procedural knowledge**

When the easy and difficult cases of procedural knowledge were combined, AAS participants had Cronbach’s alphas of .609 in Time 1, .415 in Time 2, and .687 in Time 3. NAS participants had Cronbach’s coefficients of .464 in Time 1, .310 in Time 2, and .551 in Time 3. This showed moderate reliability of the scale for participants of both studies. When assessed together, Cronbach’s alpha was .513 in Time 1, .537 in Time 2, and .666 in Time 3 for the 32-items assessed across all participants, indicating a similar moderate reliability overall for the scale across time.

**Other findings**

**Simulator Users**

At the end of the first intervention period, all 18 simulator users (AAS and NAS) were asked to list five positive and negative aspects of the simulator. These responses were obtained via an open-ended survey. A factor analysis was coded and analysed. Across all simulator users, the most identified positive aspects of the simulator was that it was easy to use (25%), had good access provided (11.11%), and it enhanced their learning (9.72%). They also found the simulator realistic (8.33%) and provided many options (8.33%). Other positive aspects included that it was practical, fun, relatively bug-free, had a good feedback system and increased their confidence.
Negative aspects centered on perceived bugs within the simulator (24.19%) and a reported poor feedback mechanism (17.74%). This was followed by not being to hear the patients’ replies; 14.52%) and the lack of control over the options on the console (12.9%). Other minor aspects reported include the realism of the virtual aspects, impractical use of the simulator, a lack of instructions given and the time taken to complete each case.

Positive aspects

AAS simulator participants reported that the software was easy to use (25.71%), the simulator had good access (17.14%), that they learnt a lot (11.43%), that the simulation was practical for their learning needs (8.57%) and was relatively bug free (8.57%).

NAS participants also found the simulator easy to use (24.32%). They found the virtual patients realistic (13.51%), and enjoyed the layout of the software (10.81%) and options available to them (10.81%) for speech audiometry testing. They also reported that they learnt a lot from the simulator (8.11%) and that it was fun (8.11%).

Negative aspects

AAS participants reported that the most negative aspects were the number of bugs (20%) and the lack of options available to them on the console (20%). They also desired to hear the patient responses (16.67%) and wanted better feedback systems for each case (16.67%).

NAS participants reported similar concerns about bugs (28.13%) and desired a better feedback system (18.75%). They, too, requested to hear patient responses (12.5%). Finally, they preferred clearer instructions for use (9.38%) and more realistic learning methods (9.38%).
Discussion

This section discusses the findings for both the expert and novice students with respect to the hypotheses. Limitations of this study are provided, and theoretical and practical implications of this study are raised. Suggestions for future research are made to conclude.

Advanced Audiology Students (AAS)

Hypothesis 1: Training transfer was found to increase for simulator users more than non-simulator users at the end of the intervention period. This increase was significant amongst simulator users. This lends support to the suggestion that simulator use can lead to good transference of learnt skills to actual clinical situations. Virtual patient use was in this case, shown to have transference into clinical situations (Issenberg et al., 1999; Sitzmann, 2011). The non-simulator group also increased post-intervention, but did not reach significance. A clear difference between groups was not found, partially due to the small associated effect sizes. However, there is enough evidence to suggest that use of the virtual patient simulator in addition to course material was useful in increasing training transfer skills.

Hypothesis 2: Declarative knowledge also increased for both groups post-intervention. This increase was not statistically significant for both groups, and may be due to the small sample sizes. This may also mean that the use of provided materials was adequate to increase declarative knowledge, and simulator use did not provide any significant enhancement in learning over this area for AAS participants. Consistent with previous research, the ability of students to verbalise pertinent facts and/or processes increased for all users regardless of simulator intervention (Herz and Schultz, 1999; Sitzmann, 2011).
Hypothesis 3: A larger increase in procedural knowledge was expected from simulator users over non-simulator users due to the constancy of use and inherent reinforcement of the stepwise clinical reasoning process (Cook & Triola, 2009). However, the opposite effect was found. Both groups had a decrease in procedural knowledge post-intervention. This decrease was more marked in the non-simulator group than in the simulator group, although this regression was not statistically significant. This unexpected finding may be a result of a more difficult test at Time 2. The procedural knowledge test at Time 2 was the least reliable (Cronbach’s alpha=.405), which may have affected the resultant scores for all participants. It appears that in this case procedural knowledge increased at the expense of declarative knowledge. Reasons provided for this decrease in procedural knowledge can include the priority given to learning declarative knowledge whilst forming new procedural knowledge, and the cumulative effect of increases in both areas may not be apparent till a later point in time (Ashby & Crossley, 2010; Herz and Schultz, 1999; Smedley & Sutton, 2007). This will be revealed by the retention knowledge measure.

Hypothesis 4: Use of the simulation was expected to reinforce and enhance learnt knowledge (Sitzmann, 2011). Training transfer was retained for both groups. It is noteworthy that for the four weeks post-intervention, simulator users regressed in their training transfer score whilst non-simulator users increased. Crucially, both of these changes were not statistically significant.

A similar pattern was found for Declarative knowledge. However, simulator users failed to retain declarative knowledge and significantly regressed. This unexpected finding of a significant decrease in simulator users’ declarative knowledge can be associated with several explanations. Firstly, the raw score differences for declarative knowledge is small for AAS
participants. Both users achieved very high scores after intervention. Although it appears that simulator users have significantly declined, it should be considered that they have gone from an excellent declarative knowledge score to a good declarative score, which is in par with the non-simulator group. Secondly, simulators were not allowed to use the simulator after intervention, which meant that they had to now learn the material solely on traditional learning methods. This essentially means that simulator users have to find new ways to recall learnt information.

Procedural knowledge recovered at Time 3 to a similar level to the baseline measure. Although the retention measure technically defines this as knowledge retained, it is more accurate to describe it as no overall change in procedural knowledge retention. However, it should be noted that simulator users had the most stability in procedural knowledge with time, with the non-simulator group declining sharply in Time 2 before recovering in Time 3.

It can be seen that for the three measures, given some time, non-simulator users are able to match simulator users in performance. Time 3 scores for both groups are closely matched across the three measures. This implicitly demonstrates that simulator users gain an advantage in learning the requisite skills quicker, and non-simulator users are still able to attain the same level of proficiency with traditional learning methods but over a longer learning period.

**Novice Audiology Students (NAS)**

In a similar manner, the three hypotheses applicable to novice students are discussed.

*Hypothesis 2:* This increase was found to be statistically significant for the simulator group but not for the non-simulator group. Therefore, this demonstrates the expected finding that simulator users would increase by more than non-simulator users. The robustness of the simulator to provide this added increase amongst novice students should be highlighted. The use
of a virtual patient simulator to increase declarative knowledge in novice students may in fact be warranted.

_Hypothesis 3:_ However, both groups had a marked decrease in procedural knowledge post-intervention. This decline was significant for both groups of users. The weak reliability of the procedural knowledge measure at time 2 (Cronbach’s alpha = .310) may partially explain this decline. This decrease was a closely followed pattern for both groups, which indicates that both groups found the test difficult and did poorly in them. This decrease in procedural knowledge combined with the increased in declarative knowledge lends further support to the notion that procedural knowledge can increase at the expense of declarative knowledge, and vice versa (Ashby & Crossley, 2010; Herz and Schultz, 1999). There was no observed difference between the two groups, indicating no simulator benefit or influence to procedural knowledge amongst novice users.

_Hypothesis 4:_ Declarative knowledge was well retained by both groups. As was the case for AAS participants, non-simulator users continued to improve after intervention whilst the simulator users marginally regressed. These patterns were not statistically significant. The patterns of results at Time 3 indicate that non-simulator users were able to catch up with simulator users over time. Simulator users were mildly disadvantaged once access to the simulator was withdrawn.

Procedural knowledge recovered significantly for both groups four weeks post-intervention, and both groups finished closely. Although Time 3 levels were significantly better than at Time 2, they were still poor when compared to the baseline. All NAS participants were assessed immediately following the initial lecture. Since some practice case examples were
covered, it may have resulted in higher procedural knowledge levels immediately following the lecture. These students were not exposed to speech audiometry outside of the study over the six-week period (in contrast to senior students), and hence it is reasonable to suggest that they quickly forgot procedural aspects of testing. However, it is disappointing that simulator use was unable to provide the necessary increase and retention of this knowledge area beyond the use of provided learning materials, and is inconsistent with previous research (Sitzmann, 2011).

**Other Findings: Simulator Users**

Based on the feedback provided by simulator users, assumptions about the simulator’s realism and the strengths of its current feedback systems were unable to be sustained. There was also some conflict in provided responses from simulator users. For example, there were variable opinions about the number of software bugs present depending on the individual user.

It is encouraging that users found the simulator easy to use and recognised that full access to the simulator was important to them. This is consistent with the implemented design theories (I.e., Issenberg et al., 2005; Sitzmann, 2011). Reports that users felt that the simulator was realistic, and that its use enhanced their learning, indicates achievement of the implied goals of utilising a virtual patient simulator design. There were also negative elements identified from this iteration of the simulator. Common areas identified were software bugs within the simulator, a poor feedback mechanism, an inability to hear the patients’ replies, and an inability to finely control options on the display console. These should arguably be regarded as positive elements, as these areas can be fixed with a future iteration of the simulator, and therefore, can easily be remedied. Overall, this is encouraging with regards to the simulator design.
Advanced audiology students in particular reported very strong areas of success with regards to the ease of using the simulator, experiencing good access, and finding it practical and useful for their learning. Their familiarity with clinical speech audiology meant that they found the software limited in the options available to them. The patients provided no auditory response, which would have been inadequate for these students. They also showed that it is unwise to rely on simulator users to refer to course notes to self-correct. Consequently, most users desired better feedback systems built into the simulator. As previously discussed, all this can be corrected.

Novice audiology students also found the simulator easy to use. Their experience was that the virtual patients were realistic and the software console was well designed with many available options. Their experience of the simulator was one of fun learning. Negative aspects identified by novice students were similar to advanced students with concerns about bugs, receiving inadequate feedback and desiring to hear patients’ responses. They also preferred clearer instructions for use, which is unsurprising given that speech audiology is new to them. Some novice students desired more realistic learning methods such as standardised patient testing. It is important to consider that these students who have not physically observed or performed speech audiology in a clinical setting still found the simulator fun and realistic. One suggestion is to use the simulator as an introduction to clinical audiology either as part of the undergraduate course or at the beginning of the Master of Audiology programme.

At the time of writing, suggestions made by simulator users were being incorporated into a future iteration of the simulator. Female and male voices have been recorded for virtual patient responses. Feedback systems, such as those suggested by Issenberg et al. (2005), Huwendiek et al., (2009) and Shute (2007) are currently being investigated and are likely to be implemented.
Alterations are being made to provide more options without altering the well-received design and layout of the console. Finally, most of the identified software bugs have been removed.

**Limitations**

There are several limitations that need to be considered when interpreting the results of this study. These limitations are present for the general research, the speech audiometry virtual patient simulator design and the overall goal of simulator use.

The sample size was inadequate to obtain the desired main and interaction effect sizes of within-between groups necessary for meaningful interpretation. A power analysis was run with the program G*Power (Faul, Erdfelder, Lang & Buchner, 2007) to ascertain the required sample size for the three stage repeated measures analysis. If only two stages were being evaluated (e.g. Pre- versus post-intervention scores, excluding retention), a large effect size of .25 (Cohen, 1988), stipulated power of .80 and significance of .05 required a sample of 34 participants. If a realistic moderate effect size of .100 is expected (Cohen, 1988), then to achieve good statistical power of .80 and significance of .05, a sample size of 164 and 200 participants was required for a three and two stage repeated measure respectively. With a final cohort of 18 participants per study, it meant that achieving meaningful interpretation values for generalisability was severely limited. The target recipients of this research was audiology students, who were limited in enrolment numbers at the University of Canterbury. This reinforces one of the aims of this research in trying to find new ways to train more audiology students. To presently incorporate more students into this study design will require having to include students from the University of Auckland, or including previous and future years of audiology students (a longitudinal analysis). Including these students would also mean altering the study design or limiting the number of dependent variables tested. For novice students, the participation restriction of
completion of the undergraduate Introduction to Audiology course is still required. Besides the previous suggestions for increasing participant numbers, targeting students at the start of the academic year may prove to be more beneficial. The nature of this Master’s research meant that implementation of the study was over the busiest times of the academic year. It is believed that this, more than any other factor, limited the number of novice participants. However, it should be noted that of the students who participated, a high completion rate was observed (94%). Mortality is sometimes cited as a possible drawback of this study design, but this was not the case for this study. It was more difficult to attract participants without a vested interest to participate in the study; but of those who did participate, nearly all completed the study. One participant who dropped out did so due to workload issues, while the other did not return for the final assessment without notification. This may indicate that the novice students who participated over the six weeks were dedicated participants, who best reflect intended recipients.

The procedural knowledge component also produced unusual results. The pattern of results was observed for all participants, where scores at Time 2 were the poorest. Reliability estimates via Cronbach’s Alpha revealed weak coefficients (.415) for advanced and novice (.310) students at Time 2. Even by combining both studies, reliability estimates for all participants was .513 in Time 1, .537 in Time 2, and .666 in Time 3, indicating moderate-to-weak reliability for Times 1 and 2, and only moderate reliability for Time 3 (Pallant, 2005). This means that each assessment of procedural knowledge was always going to produce different results, irrespective of the participant. The case based design of the procedural knowledge instrument meant that differences were intended, and the provision of an easier and more difficult case should have alleviated this problem. This was not observed given the reliability scores. Due to the unusual dip in scores in Time 2, a possible interpretation might be that both
cases used at Time 2 were too difficult for participants. This issue was not evident on the other measures of training transfer, declarative or retained knowledge.

The length of simulator intervention may have been inadequate for participants. Two weeks was set as the intervention period for theoretical and practical reasons, which was associated with improvements, of which two were significant. However, the use of the simulators may have been inadequate to cause permanent changes, as was demonstrated for declarative knowledge. Despite retaining their post-intervention knowledge, most simulator users from both studies showed subtle declines in all areas of knowledge four weeks post-intervention. This may indicate that longer intervention periods or continued use of the simulator is warranted for knowledge increase and retention. A third suggestion relates to the simulator users, who may not have depended on supplementary provided material due to their dependence on the simulator. This should be considered on a theoretical standpoint, and will be discussed in the next section. Finally, simulator design flaws may distract simulator users, which may indirectly lead to poor user experience and engagement (Cook & Triola, 2009).

Firstly, feedback provision was a challenge faced by simulator users. Numerous negative comments were received regarding the lack of feedback at the end of each case. The correct answer was displayed at the end of each case, but it did not provide information or explanations about the correct answer. More guidance needed to be provided in this area. It was assumed that the level of feedback provided was adequate since its intention was to direct students to refer to their supplementary provided material to self-correct, thus providing multiple learning strategies for students (Issenberg et al., 2005; Huwendieck et al., 2009). However, this assumption did not hold, and participants still preferred an in-built feedback mechanism. Current research (Cook & Triola, 2009; Shute, 2007) indicate that better feedback models are indeed warranted.
Secondly, software bugs were prevalent during use. Students who experienced software bugs reported them. Some of these referred to an inability to click certain buttons, software freezes and display issues. In one case, the male patient appeared as a female patient. Although these bugs are relatively minor, they still affect the users’ experience of the simulator. Fortunately, overall feedback affirmed that many found the simulator realistic, easy to use and considerably bug-free. However, software bugs need to be constantly identified and removed.

Thirdly, there was no sound provided for patients’ responses. Phonetic scoring in this study was based on the displayed responses of the patient on-screen. Genuine clinical situations require phonetic scoring of each patient’s response. The main reason this feature was not included in this version of the study was the high cost and time requirements to produce this element. Skills in phonetic scoring were still well developed when assessed by the declarative knowledge and training transfer instruments, and these elements were not found to be internally unreliable during analysis. As with the two previous limitations listed for the simulator, this limitation was being rectified in the simulator at the time of writing.

Finally, it should be noted that even if highly significant results were found through virtual patient simulator use, although it may alleviate difficulties in training areas and provide a better training tool, it would by no means indicate that students become better, more empathetic and personable clinicians. However, it may mean that their clinical reasoning may be heightened, their knowledge base both broadened and more specific, and they may have more experience with rarer cases. Combined with the clinician’s personal skills, this has the potential to effectively train high quality audiology clinicians.
Theoretical Implications

Despite the limitations, the study still remains meritorious as it contributes a further pattern of results obtained from audiology students. The dearth of literature in virtual patient simulator research receives new contributions through these audiology students. A major point of difference with this research is the well-designed, high-fidelity virtual patient simulator within audiology. This software engages the user to more actively participate in the learning process. This research is also reflective of a small but important subset of New Zealand audiology students. However, caveats ensue to prevent over-generalisation of these research findings.

There is continued support for the ABC theory frequently applied to simulator learning. Affective elements pertaining to virtual patient simulation use were not investigated in this study as these have been explored in great detail in recent years (Wilson et al., 2010; Deladisma et al., 2008; Sitzmann, 2011). With regards to behavioural elements, training transfer was significantly improved at the end of the intervention period through the use of the simulator. This agrees with Sitzmann’s (2011) findings, and indicates that the simulator may be useful in honing a student’s skills for application in a real clinical situations. However, procedural knowledge declined with intervention in both studies, and only recovered four weeks post-intervention. At first glance, it appears to conflict with Sitzmann’s (2011) work. This may not be entirely true. Limitations of the procedural knowledge scale have already been discussed. If the findings are in fact reflective of the true nature of results, it is also consistent with previous research. Decreases in procedural knowledge may be the cost of gains in declarative knowledge (Ashby & Crossley, 2010; Smedley & Sutton, 2007). Declarative knowledge may decline as the participant re-learns new behavioural skills, and during this process of increasing declarative knowledge, participants become cognitively conscious of their actions (Smedley & Sutton, 2007). As a result, procedural
knowledge becomes more adaptable to change. Thus, procedural knowledge may be temporarily negatively affected during the formation of declarative knowledge, but may recover and exceed initial levels some time after intervention. This is arguably the observation for knowledge retention, particularly amongst novice students, whose procedural knowledge significantly improved (albeit to marginally better than baseline levels) four weeks post-intervention. Being novice students, they were most capable of representing the greatest gains in any area, and coincidentally procedural knowledge gains were observable for them at Time 3.

Cognitive elements of declarative knowledge and retained knowledge also showed good support for previous findings. Declarative knowledge increases were found amongst all participants, which were expected based on previous research (Randell et al., 2007; Ifenthaler, 2010; Bell & Kozlowski, 2008; Schneider & Stern, 2010). This increase was significant only for novice students who used the simulator. Given that this was the only group where a significant change was observed, it indicates that perhaps the simulator can be used to enhance declarative knowledge at the beginning of the Master of Audiology course. If students need to be brought up to speed with speech audiometry, it is probable that simulator use coupled with coursework may quickly achieve this goal.

Retained knowledge was more broadly defined in this study. A more inclusive definition of retained knowledge applied to all knowledge areas being assessed, and not just declarative knowledge (Botezatu, 2010b; Kerfoot & Brotschi, 2009; Sitzmann, 2011). Knowledge was retained generally well across the study with two exceptions, and potential simulator benefit can be seen. The first was that procedural knowledge at Time 3 improved for both groups. This unusual pattern has already been discussed. The other exception was that declarative knowledge for expert student simulator users was not retained at the end of the study. This pattern was
observed for most simulator users (expert and novice students), and was found to be a significant decline for this group with declarative knowledge. This unusual finding may be potentially caused by the inverse relationship between procedural and declarative knowledge. Although this is possible, this is not likely since knowledge retention should be somewhat independent of these changes. It is more likely that simulator users were more reliant on the simulator during intervention. After intervention, when users were no longer able to use the simulator, increases first observed immediately post-intervention were not as strong. By the final assessment four weeks later, simulator users would have had to develop a second strategy for learning not dependent on simulator use (i.e. akin to the non-simulator users) (Botezatu, 2010b). This modification of learning strategies may have contributed to the decline in their scores at Time 3. From a practical standpoint, knowledge retention is linked to how much continuous exposure there is with the simulator after intervention. The theoretical basis of this assertion will need to be better explored in future research; however, practical considerations of this suggestion will still be made in the next section.

There were also differences in outcomes observed between novice and expert users. This indicates that the user’s year of study should be considered in virtual patient simulation research. The timing of virtual patient simulator introduction within a learning context is crucial with regards to the targeted learning outcomes. Simulator use should not be without clear goals, and careful consideration of the intended outcome may warrant its inclusion and time of introduction (Botezatu, 2010b). Expert users may look to practice core principles and consolidate previously learnt areas, whilst novice users would be simply getting adjusted to a new learning system (Dawson & Gould, 2007).
The relative success and strong feedback provided by simulator users demonstrate that the virtual patient simulator was already of a good standard at this point. Design elements obtained from the literature review (e.g. Issenberg et al., 2005; Cook & Triola, 2009; Huwendiek et al., 2009) were implemented in the design. Of the ten recommendations made in the literature review, all were included at various levels. Participants’ experience of the simulator and its design was largely positive with the exception of resolvable software bugs. Therefore, support and encouragement of the virtual patient essentials provided by Issenberg et al. (2005), Huwendiek et al. (2009), and Cook & Triola (2009) should be endorsed for virtual patient simulator design. The main drawback previously discussed was focused on feedback.

Positively identified elements included users complete access to the simulator during the time of intervention. Cases were reproducible and standardised learning activities, and were built to be realistic and responsive to student’s needs. A wide range of clinical situations were presented within a controlled learning environment. Users were freely able to make, detect and correct mistakes without adverse consequences. Learning outcomes and goals were both clear and implicit. Seven of the nine tasks of speech audiometry could be easily performed of which clear outcomes and goals could be set. The display of results at the conclusion of each case only after half-peak levels and consistency with the patient’s audiogram had been established, allowing the user to implicitly determine if they had met learning goals. To this end, this was well achieved, and was reflected by the students.
Practical implications

Meeting the challenge of finding more effective training methods for audiology students will require more than virtual patient simulator use. This study demonstrated that the virtual patient simulator is useful, but many moderating factors will hamper its inclusion and widespread introduction at this point.

There is enough evidence to suggest that the simulator should be included in the course curriculum. Simulator use can have many of the benefits discussed for expert and novice students (Dawson & Gould, 2007). Given the other learning commitments faced by students, it is unintended that the simulator becomes simply another tool that they are expected to be proficient with. Learning outcomes with the simulator need to be constantly reviewed to ensure that any simulator use is targeted towards specific learning outcomes (Botezatu, 2010b; Issenberg et al., 2005). It is important to consider that in this study, and as recommended by other research (Issenberg et al., 2005; Sitzmann, 2011), the simulator cannot be the sole medium of instruction. The simulator use must continue to be coupled with course notes, lectures and clinical experiences. This was the case for audiology students in this study. Design principles incorporated into this simulator such as complete access to students need to be maintained if the simulator is to be introduced.

Non-simulator users were also able to “catch up” with any significant improvement observed in the simulator group by the end of the study. By implication, this means that either the simulator group regressed, or the non-simulator group improved. Improvements by the non-simulator group indicate that the simulator is able to accelerate learning. If this is true, then the simulator should be introduced near the start of a course. If the improvement was due to regression of simulator users’ non-use, then careful management of simulator use is essential.
Targeted and regular use of the simulator may in fact enhance retained knowledge. However, this finding was inconclusive. Therefore, simulator use must be provided for an extended period of time to provide genuine and deep learning to take place. Students must be reminded that other forms of the course material are just as important for their learning. This will prevent unrealistic expectations about simulator use to be formed. Clear learning outcomes utilising the speech audiometry tasks identified by the literature review would also be useful. It is also recommended that training transfer and declarative knowledge instruments developed for this study are used for clinical assessment of students.

Suggestions by users for improving feedback systems, removing software bugs and providing patient sound output have been discussed earlier. More explicit instruction needs to be provided for users to self-correct more effectively. Design limitations and the early stages of the speech audiometry simulator mean that more advanced feedback systems could not be introduced for the purposes of this study. However, given the overwhelming request and better features required, it is imperative that it is modeled in any future iterations of the speech audiometry simulator. Some of these changes may include having specified reasons for incorrect answers, and suggestions provided at the end of each case. Hints may also appear on-screen during the case if clear mistakes are made or if steps are missed. Formulae required for presenting speech stimuli and for masking should be accessible with the click of a button. If a preference is still to include multiple learning strategies, then the simulator may ask them to refer to certain parts of their notes to clarify confusing concepts. More research is required to validate these components, and determine which mix of feedback components best enhances and sustains learning.
Software bugs will need to be removed to ensure a smooth and clear operation of the simulator. This is not limited to virtual patient simulators, but any computer software. Alterations are being made to unlock a few options without altering the fancied design and layout console options, and most of the identified software bugs have been removed. Cases from previous work on the pure-tone audiometry and case history components are being combined into complete cases, where users can either choose to practice individual components or perform the complete audiometric test battery. The freedom to do this in their own time without having to enlist the help of a volunteer participant can empower these students to pursue learning at their own pace. Provided that all students experience use of the virtual patient simulator, it may even be possible to teach or assess multiple students at once using any required virtual patient case. If the cases are compiled with validated information from the previously designed elements of pure-tone audiometry testing and case history taking, then complete cases for practice can be established for students. Finally, the virtual patient responses can still be made to be even more realistic for speech audiometry testing. Recorded voices are required to mimic real patients. 1200 words are required for each patient in the software and hence was not introduced at this point. However, at the time of writing, these essential elements are currently being evaluated. Male and female voices were being recorded and were to be added to a future iteration of the simulator.

Audiology training in New Zealand can utilise this software using the recommendations provided above. This can apply to other Universities as well as new migrant audiologists who wish to learn the New Zealand Audiological Society protocols. Although this simulator has been designed and developed for the New Zealand context, its validation means that cross-principles can be applied in designing for other countries’ requirements. Consideration should be made for multilingual audiological simulators currently being designed (Kompis et al., 2012). Broader
application of the simulator should also be considered. Besides training purposes, the virtual patient simulator can potentially be used as a pre-procedural planning and rehearsal tool (Willaert et al., 2012). Virtual patient simulator use can be combined with learning methods that assist the development of genuine personal relationship building skills. This will overcome any technical limitations of computer-based virtual patients with meaningful training that can be genuinely applied to clinical situations. This, combined with practical behavioural knowledge, can be used to train highly competent, effective, personable and empathetic clinicians.

**Future research**

This study has encouraged the need for further research in this area. If audiology is going to continue to implement virtual patient simulation, research into reproducible learning goals are warranted. Enhancing clinical reasoning skills (Cook & Triola, 2009) is one of the main goals of virtual patient simulation use. Better measures, more suited to assessing clinical reasoning in students, need to be developed. If a well-designed, validated and reliable instrument is available for use, then research should be undertaken if audiology students can demonstrate clinical reasoning benefit through the use of virtual patient simulators. The simulator may be more useful when it incorporates a better feedback and support system with no software bugs. It is then that more in depth analysis can be undertaken to identify holistic learning areas.

Future research should not be limited to simply increasing sample sizes. However, the use of broader sample sizes achieved by incorporating audiology students from around New Zealand (and possibly Australia) should be considered. This could allow more representative sampling characteristics to be observed in an experimental design.
As the previous sections have thoroughly established, direct feedback mechanisms are essential for student learning. This was found to be somewhat lacking in this study. More comprehensive feedback tools should be built into the virtual patient simulator. This requires further research to establish optimal feedback delivery and suitability for audiology students.

Finally, a review of other learning solutions to overcome the difficulties in training New Zealand audiologists can be explored. Wider solutions, including industry incentives for training clinical educators, or restructuring clinical placements can be further explored. Clinical placements supervisors should be encouraged to mentor students on in a way that ensures that the currency of their own knowledge is preserved, whilst providing the student with practical, useful and hands-on learning experiences to improve their knowledge base and clients before interacting with real clients. However, the way to overcome these problems remains in the domain of more innovative research.

Conclusions

One of the ways to overcome the problem of the shortage of audiologists in New Zealand is by improving the training effectiveness of audiology students. This can be achieved through the use of a virtual patient simulator (VPS). This research explored if a VPS designed for the New Zealand speech audiometry protocol was able to achieve required learning outcomes.

In summary, the development of a VPS for New Zealand purposes was found to be useful for some learning areas for students. Key findings were that expert students who used the simulators were found to improve their training transfer skills, whilst novice students had significant improvements in their declarative knowledge. This difference was observed after only
two weeks intensive VPS use combined with other learning material. Procedural knowledge was found to be inconclusive after intervention. Overall, knowledge was well retained for all participants in this study. Any differences between groups usually recovered six weeks post-intervention. There is evidence to suggest that a longer period of intervention may be useful for knowledge retention, and that use of the simulator may quicken the learning process of audiological clinical practices (in this case, speech audiometry).

Overall, simulator users found the virtual patient software easy to use and easy to access as required. They found that it enhanced their knowledge, was realistic and had many usable options for testing purposes. These are positive elements crucial to continued use of any computer-based learning simulators, and are encouraging in the validation of the speech audiometry component of the simulator. Suggestions and negative aspects were also noted and will be implemented in future iterations of the simulator. These included providing more detailed feedback to users during and after testing, as well as providing audio output so patient responses can be heard. Finally, there is a call to reduce software bugs. These can be easily rectified, and were being implemented at the time of writing.

Theoretical contributions to research have been suggested, and future directions for simulator design and use as part of a holistic education have been made. This will mean incorporating a mixture or ordinary learning materials, clinical placements and regular, targeted simulator use. This will have to be managed within the time constraints of each student. There is evidence to suggest that both expert and new students will benefit from simulator use, and simulator introduction and learning outcomes should be adjusted to the recipient students.
Future research directions include assessing using a complete audiometry simulator with a range of challenging cases, as well as continually identifying new ways of overcoming the ongoing difficulty of training sufficient audiologists in New Zealand. This virtual patient simulator remains a viable and pragmatic medium-term solution for this pertinent issue.
References


### Appendices

#### Appendix 1: Training Transfer Marking Schedule for Standardised Patients

<table>
<thead>
<tr>
<th>Item</th>
<th>Potential Marks</th>
<th>Awarded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Better ear</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Worse ear</td>
<td></td>
</tr>
<tr>
<td>Instructions to clients (-1 for missing instruction)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Reading audiogram and otsocopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start with better ear</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Start with correct transducers</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Calibration</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Correctly done</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Presentations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pl Max</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start at appropriate presentation level</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Correct phonetic scoring (4 marks, -½ for each mistake)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Record mistakes (If req., if not full marks)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Correct sum (1 mark)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Score recorded on audiogram (-1 if incorrect symbols, -1 if wrong placement)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Masking required (excluded if unneeded)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Correct masking level (excluded if unneeded)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Correct stimuli (-1 for incorrect stimuli)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Slope (2nd point)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain higher level if necessary, if not for appropriate slope level</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Correct phonetic scoring (4 marks, -½ for each mistake)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Record mistakes (If req., if not full marks) (1 mark)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Correct sum (1 mark)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Score recorded on audiogram (-1 if incorrect symbols, -1 if wrong placement)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Masking required (excluded if unneeded)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Correct masking level (excluded if unneeded)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Correct stimuli (-1 for incorrect stimuli)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Half-peak level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate level</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Correct phonetic scoring (4 marks, -½ for each mistake) | 4 | 4
Record mistakes (If req., if not full marks) (1 mark) | 1 | 1
Correct sum (1 mark) | 1 | 1
Score recorded on audiogram (-1 if incorrect symbols, -1 if wrong placement) | 2 | 2
Masking required (excluded if unneeded) | 4 | 4
Correct masking level (excluded if unneeded) | 6 | 6
Correct stimuli (-1 for incorrect stimuli) | 2 | 2

**General**
Pausing as required | 1 | 1

**Estimation**
PI Max percentage | 2 | 2
Half-peak level | 2 | 2

Consistency with audiogram | 1 | 1
Reason given | 2 | 2

Feeding back to client (-1 for not attempting to explain purpose of speech audiometry) | 2 |

**Efficiency**
At least 3 points per ear | 1 | 1
5 or less points per ear | 1 | 1
Less than or equal to 10 minutes | 1 |
Less than or equal to 15 minutes | 1 |
Less than or equal to 20 minutes | 1 |

GRAND TOTAL (OUT OF 165) | 89 | 76

**Notes:**

1. The items highlighted in yellow indicate the items excluded from the training transfer measure during analysis.
2. The items highlighted in green indicate the items not score in each of the 3 Standardised Patient cases.
3. Each case was marked on 59-of the above items (i.e., the non-highlighted items);
4. Out of a total score of 114 points.
Appendix 2A: Procedural knowledge questions at Time 1

Easier case

This 35 year-old woman has come in to have her hearing checked as part of a job interview.

There are no complains of balance problems, tinnitus, or other health concerns. Here is her audiogram.

1. Which ear should you test first? (half mark to be given if corresponding to 1)
   a. Right ear
   b. Left ear
   c. Either ear

2. Here are the otoscopy images for her left and right ears respectively:
Considering them both together, which of the following transducers is most recommended?
   a. Supra-Aural headphones
   b. Bone-conductor headphones
   c. Insert headphones
   d. Soundfield

3. Is the hearing loss configuration for her right ear flat or sloping?
   a. Flat
   b. Sloping
   c. Rising

4. Based on 3), what is the initial presentation level for speech in her right ear?
   a. 50 dB
   b. 25 dB
   c. 45 dB
   d. 40 dB

5. Is the hearing loss configuration for her left ear flat or sloping?
   a. Sloping
   b. Rising
   c. Flat

6. Based on 5), what is the initial presentation level for speech in her left ear?
   a. 45 dB
   b. 40 dB
   c. 25 dB
   d. 35 dB

7. Based on your answers for questions 2-6, would you need to put masking into the better ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

8. How much masking is required for the initial presentation levels from questions 2-6?
   a. No masking is required
   b. 0 dB HL
   c. -5 dB HL
   d. 5 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?
    a. No masking is required
    b. 0 dB HL
    c. -5 dB HL
    d. 5 dB HL
11. This is her speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.

![Speech Audiometry Graph]

What is the PI Max of her Right ear?
- a. 100%
- b. 93%
- c. 83%
- d. 95%

12. What is the Half-peak Level for her Right ear?
- a. 10 dB HL
- b. 12 dB HL
- c. 15 dB HL
- d. 50 dB HL

13. Are the results for her right ear consistent with the audiogram?
- a. Yes because the HPL is within 15 dB of the 1 kHz threshold
- b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
- c. No because the HPL is more than 15 dB away from the 1 kHz threshold
- d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz

14. What is the PI Max of her Left ear?
- a. 93%
- b. 83%
- c. 97%
- d. 100%

15. What is the Half-peak Level for her Left ear?
- a. 8 dB HL
- b. 10 dB HL
- c. 12 dB HL
- d. 50 dB HL

16. Are the results for her left ear consistent with the audiogram?
- a. Yes because the HPL is within 15 dB of the 1 kHz threshold
- b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
- c. No because the HPL is more than 15 dB away from the 1 kHz threshold
- d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
**Difficult case**

This 27 year-old man received a traumatic injury to the right ear when playing around with a chopstick. The ear drum is healing but the hearing still “feels down” in the right ear. He wants to have it checked and see what can be done.

1. Which ear should you test first?
   a. Right ear
   b. Left ear
   c. Either ear

2. Here are the otoscopy images for his right and left ears respectively.

![Otoscopes](image-url)
Considering them both together, which of the following transducers is most recommended?

a. Supra-Aural headphones  
b. Bone-conductor headphones  
c. Insert headphones  
d. Soundfield

3. Is the hearing loss configuration for his right ear flat or sloping?

a. Flat  
b. Sloping  
c. Rising

4. Based on 3), what is the initial presentation level for speech in his right ear?

a. 50 dB  
b. 65 dB  
c. 80 dB  
d. 90 dB

5. Is the hearing loss configuration for his left ear flat or sloping?

a. Sloping  
b. Rising  
c. Flat

6. Based on 5), what is the initial presentation level for speech in his left ear?

a. 50 dB  
b. 30 dB  
c. 20 dB  
d. 45 dB

7. Based on your answers for questions 2-6, do you need to put masking into the better ear? (mark accordingly)

a. Yes, because interaural attenuation is going to affect the other ear  
b. No, because sound will not cross over to the other ear  
c. Yes, because we have normal hearing in the better ear  
d. No, because there is no air bone gap

8. How much masking is required for the initial presentation levels from questions 2-6?

a. No masking is required  
b. 60 dB HL  
c. 40 dB HL  
d. 50 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)

a. Yes, because interaural attenuation is going to affect the other ear  
b. No, because sound will not cross over to the other ear  
c. Yes, because masking is required for flat hearing losses  
d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?

a. No masking is required  
b. 0 dB HL  
c. -5 dB HL  
d. 5 dB HL

11. This is his speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.
What is the PI Max of his Right ear?
   a. 97%
   b. 93%
   c. 100%
   d. 95%

12. What is the Half-peak Level for his Right ear?
   a. 56 dB HL
   b. 50 dB HL
   c. 12 dB HL
   d. 80 dB HL

13. Are the results for his right ear consistent with the audiogram?
   a. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
   b. Yes because the HPL is within 15 dB of the 1 kHz threshold
   c. No because the HPL is more than 15 dB away from the 1 kHz threshold
   d. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz

14. What is the PI Max of his Left ear?
   a. 93%
   b. 83%
   c. 97%
   d. 100%

15. What is the Half-peak Level for his Left ear?
   a. 15 dB HL
   b. 50 dB HL
   c. 12 dB HL
   d. 10 dB HL

16. Are the results for his left ear consistent with the audiogram?
   a. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
   b. No because the HPL is more than 15 dB away from the 1 kHz threshold
   c. Yes because the HPL is within 15 dB of the 1 kHz threshold
   d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
Appendix 2B: Procedural knowledge questions at Time 2

Easier case

This 66 year-old lady has noticed that her hearing has got gradually worse over several years, and is struggling to understand speech in noisy situations. She wishes to have her hearing tested to see if hearing aids would do her any good.

1. Which ear should you test first? (half mark to be given if corresponding to 1)
   a. Right ear
   b. Left ear
   c. Either ear

2. Here are the otoscopy images for her left and right ears respectively.
Considering them both together, which of the following transducers is most recommended?

a. Supra-Aural headphones
b. Bone-conductor headphones
c. Insert headphones
d. Soundfield

3. Is the hearing loss configuration for her right ear flat or sloping?
   a. Flat
   b. Sloping
   c. Rising

4. Based on 3), what is the initial presentation level for speech in her right ear?
   a. 40 dB
   b. 50 dB
   c. 60 dB
   d. 30 dB

5. Is the hearing loss configuration for her left ear flat or sloping?
   a. Sloping
   b. Rising
   c. Flat

6. Based on 5), what is the initial presentation level for speech in her left ear?
   a. 30 dB
   b. 50 dB
   c. 20 dB
   d. 40 dB

7. Based on your answers for questions 2-6, would you need to put masking into the better ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

8. How much masking is required for the initial presentation levels from questions 2-6?
   a. No masking is required
   b. 0 dB HL
   c. 40 dB HL
   d. 10 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?
    a. No masking is required
    b. 0 dB HL
    c. -5 dB HL
    d. 5 dB HL

11. This is her speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.
What is the PI Max of her Right ear?
  a. 100%
  b. 93%
  c. 97%
  d. 90%

12. What is the Half-peak Level for her Right ear?
  a. 33 dB HL
  b. 35 dB HL
  c. 37 dB HL
  d. 30 dB HL

13. Are the results for her right ear consistent with the audiogram?
  a. Yes because the HPL is within 15 dB of the 1 kHz threshold
  b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
  c. No because the HPL is more than 15 dB away from the 1 kHz threshold
  d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz

14. What is the PI Max of her Left ear?
  a. 93%
  b. 83%
  c. 97%
  d. 100%

15. What is the Half-peak Level for her Left ear?
  a. 38 dB HL
  b. 36 dB HL
  c. 34 dB HL
  d. 32 dB HL

16. Are the results for her left ear consistent with the audiogram?
  a. Yes because the HPL is within 15 dB of the 1 kHz threshold
  b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
  c. No because the HPL is more than 15 dB away from the 1 kHz threshold
  d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
**Difficult case**

A 62 year-old woman notices that she is having more difficulty hearing with her left ear than her right. This lady regularly uses cotton buds but does not feel that it makes a difference. There is no aural fullness or facial numbness.

1. Which ear should you test first?
   a. Right ear
   b. Left ear
   c. Either ear
2. Here are the otoscopy images for her left and right ears respectively.
Considering them both together, which of the following transducers is most recommended?

a. Supra-Aural headphones  
b. Bone-conductor headphones  
c. Insert headphones  
d. Soundfield

3. Is the hearing loss configuration for her right ear flat or sloping?

a. Flat  
b. Sloping  
c. Rising

4. Based on 3), what is the initial presentation level for speech in her right ear?

a. 50 dB  
b. 60 dB  
c. 80 dB  
d. 70 dB

5. Is the hearing loss configuration for her left ear flat or sloping?

a. Sloping  
b. Rising  
c. Flat

6. Based on 5), what is the initial presentation level for speech in her left ear?

a. 60 dB  
b. 65 dB  
c. 75 dB  
d. 85 dB

7. Based on your answers for questions 2-6, do you need to put masking in the better ear? (mark accordingly)

a. Yes, because interaural attenuation is going to affect the other ear  
b. No, because sound will not cross over to the other ear  
c. Yes, because we have normal hearing in the better ear  
d. No, because there is no air bone gap

8. How much masking is required for the initial presentation from questions 2-6?

a. No masking is required  
b. 35 dB HL  
c. 45 dB HL  
d. 55 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)

a. Yes, because interaural attenuation is going to affect the other ear  
b. No, because sound will not cross over to the other ear  
c. Yes, because masking is required for flat hearing losses  
d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?

a. No masking is required  
b. 0 dB HL  
c. -5 dB HL  
d. 5 dB HL

11. This is her speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.
What is the PI Max of her Right ear?
  a. 97%
  b. 94%
  c. 100%
  d. 90%

12. What is the Half-peak Level for her Right ear?
  a. 36 dB HL
  b. 38 dB HL
  c. 33 dB HL
  d. 30 dB HL

13. Are the results for her right ear consistent with the audiogram?
  a. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
  b. Yes because the HPL is within 15 dB of the 1 kHz threshold
  c. No because the HPL is more than 15 dB away from the 1 kHz threshold
  d. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz

14. What is the PI Max of her Left ear?
  a. 93%
  b. 94%
  c. 97%
  d. 100%

15. What is the Half-peak Level for her Left ear?
  a. 60 dB HL
  b. 62 dB HL
  c. 67 dB HL
  d. 65 dB HL

16. Are the results for the left ear consistent with her audiogram?
  a. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
  b. No because the HPL is more than 15 dB away from the 1 kHz threshold
  c. Yes because the HPL is within 15 dB of the 1 kHz threshold
  d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
Appendix 2C: Procedural knowledge questions at Time 3

Easier case

This 74 year-old woman has come in to get tested following a referral by the GP, whom she saw for vertigo and nausea every time she moves her head. She knows her hearing and balance has worsened over many years, but does not experience tinnitus.

1. Which ear should you test first? (half mark to be given if corresponding to 1)
   a. Right ear
   b. Left ear
   c. Either ear

2. Here are the otoscopy images for her right and left ears respectively.
Considering them both together, which of the following transducers is most recommended?

a. Supra-Aural headphones
b. Bone-conductor headphones
c. Insert headphones
d. Soundfield

3. Is the hearing loss configuration for her right ear flat or sloping?
   a. Flat
   b. Sloping
   c. Rising

4. Based on 3), what is the initial presentation level for speech in her right ear?
   a. 65 dB
   b. 50 dB
   c. 55 dB
   d. 80 dB

5. Is the hearing loss configuration for her left ear flat or sloping?
   a. Sloping
   b. Rising
   c. Flat

6. Based on 5), what is the initial presentation level for speech in her left ear?
   a. 80 dB
   b. 50 dB
   c. 60 dB
   d. 70 dB

7. Based on your answers for questions 2-6, do you need to put masking into the better ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

8. How much masking is required for the initial presentation from questions 2-6?
   a. No masking is required
   b. 0 dB HL
   c. 40 dB HL
   d. 10 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)
   a. Yes, because interaural attenuation is going to affect the other ear
   b. No, because sound will not cross over to the other ear
   c. Yes, because masking is required for flat hearing losses
   d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?
    a. No masking is required
    b. 0 dB HL
    c. -5 dB HL
    d. 5 dB HL

11. This is her speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.
What is the PI Max of her Right ear?
   a. 100%
   b. 93%
   c. 90%
   d. 97%

12. What is the Half-peak Level for her Right ear?
   a. 57 dB HL
   b. 53 dB HL
   c. 59 dB HL
   d. 55 dB HL

13. Are the results for her right ear consistent with the audiogram?
   a. Yes because the HPL is within 15 dB of the 1 kHz threshold
   b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
   c. No because the HPL is more than 15 dB away from the 1 kHz threshold
   d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz

14. What is the PI Max of her Left ear?
   a. 93%
   b. 90%
   c. 97%
   d. 100%

15. What is the Half-peak Level for her Left ear?
   a. 60 dB HL
   b. 58 dB HL
   c. 56 dB HL
   d. 54 dB HL

16. Are the results for her left ear consistent with the audiogram?
   a. Yes because the HPL is within 15 dB of the 1 kHz threshold
   b. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
   c. No because the HPL is more than 15 dB away from the 1 kHz threshold
   d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
**Difficult case**

This 68 year-old lady has noticed a drop in hearing over many years, and is here because of a sudden attack of vertigo and nausea in the last 2 days. To rule out everything else, a hearing test is requested by the GP.

1. Which ear should you test first?
   a. Right ear
   b. Left ear
   c. Either ear
2. Here are the otoscopy images for her right and left ears respectively.
Considering them both together, which of the following transducers is most recommended?
  a. Supra-Aural headphones
  b. Bone-conductor headphones
  c. Insert headphones
  d. Soundfield

3. Is the hearing loss configuration for her right ear flat or sloping?
  a. Flat
  b. Sloping
  c. Rising

4. Based on 3), what is the initial presentation level for speech in her right ear?
  a. 75 dB
  b. 80 dB
  c. 85 dB
  d. 90 dB

5. Is the hearing loss configuration for her left ear flat or sloping?
  a. Sloping
  b. Rising
  c. Flat

6. Based on 5), what is the initial presentation level for speech in her left ear?
  a. 20 dB
  b. 25 dB
  c. 40 dB
  d. 45 dB

7. Based on your answers for questions 2-6, do you need to mask the better ear? (mark accordingly)
  a. Yes, because interaural attenuation is going to affect the other ear
  b. No, because sound will not cross over to the other ear
  c. Yes, because we have normal hearing in the better ear
  d. No, because there is no air bone gap

8. How much masking is required for the initial presentation from questions 2-6?
  a. No masking is required
  b. 50 dB HL
  c. 40 dB HL
  d. 30 dB HL

9. Based on your answers for questions 2-6, would you need to put masking into the worse ear? (mark accordingly)
  a. Yes, because interaural attenuation is going to affect the other ear
  b. No, because sound will not cross over to the other ear
  c. Yes, because masking is required for flat hearing losses
  d. No, because there is no air bone gap

10. How much masking is required for the initial presentation levels from questions 2-6?
  a. No masking is required
  b. 0 dB HL
  c. -5 dB HL
  d. 5 dB HL

11. This is her speech audiogram for both ears (Fig C). Use this information to answer questions 11-16.
What is the PI Max of her Right ear?
   a. 97%
   b. 93%
   c. 90%
   d. 100%

12. What is the Half-peak Level for her Right ear?
   a. 76 dB HL
   b. 74 dB HL
   c. 72 dB HL
   d. 70 dB HL

13. Are the results for her right ear consistent with the audiogram?
   a. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
   b. Yes because the HPL is within 15 dB of the 1 kHz threshold
   c. No because the HPL is more than 15 dB away from the 1 kHz threshold
   d. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz

14. What is the PI Max of her Left ear?
   a. 93%
   b. 90%
   c. 97%
   d. 100%

15. What is the Half-peak Level for her Left ear?
   a. 21 dB HL
   b. 23 dB HL
   c. 25 dB HL
   d. 27 dB HL

16. Are the results for her left ear consistent with the audiogram?
   a. Yes because the HPL is within 15 dB of the average loss between 2-4 kHz
   b. No because the HPL is more than 15 dB away from the 1 kHz threshold
   c. Yes because the HPL is within 15 dB of the 1 kHz threshold
   d. No because the HPL is more than 15 dB away from the average loss between 2-4 kHz
Appendix 3: Declarative knowledge questions

**Time 1**

1. What is the interaural attenuation for supra-aural headphones for speech audiometry?
2. What is the formula for initial presentation level for a flat audiogram?
3. Here is a word list. Complete the scoring and total score.

<table>
<thead>
<tr>
<th>Word</th>
<th>Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean</td>
<td>Leap</td>
<td></td>
</tr>
<tr>
<td>Hag</td>
<td>Heap</td>
<td></td>
</tr>
<tr>
<td>Bed</td>
<td>Dance</td>
<td></td>
</tr>
<tr>
<td>Sews</td>
<td>Ruse</td>
<td></td>
</tr>
<tr>
<td>Cop</td>
<td>Cot</td>
<td></td>
</tr>
<tr>
<td>Root</td>
<td>Root</td>
<td></td>
</tr>
<tr>
<td>Pick</td>
<td>Pop</td>
<td></td>
</tr>
<tr>
<td>Maim</td>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>Toss</td>
<td>Tom</td>
<td></td>
</tr>
<tr>
<td>Dial</td>
<td>Dial</td>
<td></td>
</tr>
<tr>
<td>STIMULUS</td>
<td>50 dB HL in left ear</td>
<td></td>
</tr>
<tr>
<td>MASKING</td>
<td>10 dB HL in right ear</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. What symbol is used to mark this list on the audiogram?
5. If this was the first list for this ear, would you increase or decrease the presentation level?
6. What is the formula for deciding if you need to mask?
7. What is the formula for deciding how much masking to put in?
8. What type of masking noise is used for speech audiometry?
9. How do you check if the half-peak level is consistent if the audiogram is sloping?
10. If the PI Max is 90%, what is the percentage for its half-peak level?
11. How many points should we obtain per ear?
12. What are the three purposes of speech audiometry?

**Time 2**

1. What is the interaural attenuation for insert headphones for speech audiometry?
2. What is the formula for initial presentation level for a steeply sloping audiogram?
3. Here is a word list. Complete the scoring and total score.

<table>
<thead>
<tr>
<th>Word</th>
<th>Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hide</td>
<td>Book</td>
<td></td>
</tr>
<tr>
<td>Tame</td>
<td>Tame</td>
<td></td>
</tr>
<tr>
<td>Rule</td>
<td>Fool</td>
<td></td>
</tr>
<tr>
<td>Cause</td>
<td>Poz</td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>Sick</td>
<td></td>
</tr>
<tr>
<td>Sass</td>
<td>Suss</td>
<td></td>
</tr>
<tr>
<td>Pope</td>
<td>Par</td>
<td></td>
</tr>
<tr>
<td>Don</td>
<td>Pen</td>
<td></td>
</tr>
</tbody>
</table>
Meek  | Chair  |  
Let  | Loop  |  
**STIMULUS**  | 30 dB HL in right ear  |  
**MASKING**  | No masking in left ear  |  
**TOTAL**  |  |  

4. What symbol is used to mark this list on the audiogram?
5. If this was the first list for this ear, would you increase or decrease the presentation level?
6. What is the formula for deciding if you need to mask?
7. What is the formula for deciding how much masking to put in?
8. Why do we need test at levels over 90 dB HL for speech audiometry?
9. How do you check if the half-peak level is consistent if the audiogram is flat?
10. If the Pi Max is 94%, what is the percentage for its half-peak level?
11. How many points should we obtain per ear?
12. What are the three purposes of speech audiometry?

**Time 3**

1. What is the interaural attenuation for insert headphones for speech audiometry?
2. What is the formula for initial presentation level for a flat audiogram?
3. Here is a word list. Complete the scoring and total score.

<table>
<thead>
<tr>
<th>Word</th>
<th>Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make</td>
<td>Rake</td>
<td></td>
</tr>
<tr>
<td>Laws</td>
<td>Law</td>
<td></td>
</tr>
<tr>
<td>Rice</td>
<td>Mice</td>
<td></td>
</tr>
<tr>
<td>Bell</td>
<td>Well</td>
<td></td>
</tr>
<tr>
<td>Tote</td>
<td>Toad</td>
<td></td>
</tr>
<tr>
<td>Cod</td>
<td>Hid</td>
<td></td>
</tr>
<tr>
<td>Ham</td>
<td>Beam</td>
<td></td>
</tr>
<tr>
<td>Deep</td>
<td>Meek</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>Pig</td>
<td></td>
</tr>
<tr>
<td>Soon</td>
<td>Soon</td>
<td></td>
</tr>
<tr>
<td><strong>STIMULUS</strong></td>
<td>90 dB HL in left ear</td>
<td></td>
</tr>
<tr>
<td><strong>MASKING</strong></td>
<td>0 dB HL in right ear</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. What symbol is used to mark this list on the audiogram?
5. If this was the first list for this ear, would you increase or decrease the presentation level?
6. What is the formula for deciding if you need to mask?
7. What is the formula for deciding how much masking to put in?
8. Speech audiometry uses CVC word lists. What does CVC stand for?
9. How do you check if the half-peak level is consistent if the audiogram is flat?
10. If the Pi Max is 82%, what is the percentage for its half-peak level?
11. How many wordlists should we present per ear?
12. What are the three purposes of speech audiometry?