University of Southampton

2010

Faculty of Engineering, Science and Mathematics

Institute of Sound and Vibration Research

The Inter-Rater Reliability of Categories of Auditory Performance-II (CAP)-II

By

Louise Gilmour

MSc Audiology

2010

Supervisors: Ms Julie Brinton &

Dr Carl Verschuur

Abstract

The aim of the present study was to validate a modified version of the Categories of Auditory Performance (CAP) rating scale, the CAP-II. The CAP is a hierarchy rating scale which assesses a child's functioning in everyday situations. It covers a range of auditory performance and also takes into consideration different developmental rates of children. Inter-rater reliability was assessed by participants rating benefits young children receive with cochlear implants (CI). Volunteers watched video recordings of staff describing hypothetical cases of children who wear CIs.

Children are difficult to assess using conventional methods, requiring their cooperation, thus there is need for additional assessment methods. The CAP has been validated and can provide additional information on disability reduction beyond that provided by audiological tests, as it illustrates the benefits that children now receive with CIs. The original CAP scale consists of 8 categories. There is evidence of ceiling effects as it does not address the more complicated listening skills achievable with CIs. Therefore it has been proposed that two new additional categories are added to the original scale to form the CAP-II.

Inter-rater reliability was assessed using video recordings of rehabilitative staff describing behaviour and listening skills of children with CIs. Thirty-three participants, volunteers from the University of Southampton watched 6 video recordings and assigned ratings based on how each child was doing. High inter-rater reliability was found between all participants. The percent of overall agreement (PO) was 0.76% and no significant difference was found between the ratings ($p>0.05$). There was also a strong relationship between all participants, as there was a statistically significant correlation between participants when they were compared ($p<0.05$).

Overall the results indicate that the modified CAP has good reliability. It is proven to be a relatively reliable and useful method. Specific findings about the CAP-II especially categories 7 (use of telephone with known speaker), 8 (follow group conversation in a reverberant room or where there is some interfering noise such as in a classroom or restaurant) and 9 (use of telephone with unknown speaker in unpredictable context) suggest that they need to be reordered. Categories 7 and 9 should be optional. Young children may not yet have developed these listening skills due to their age. Once these modifications are made the CAP-II could prove to be a valid tool in clinical use, as it addresses the additional benefits now achievable with CIs whilst avoiding ceiling effects. Although more research involving a larger population needs to be conducted in order to assess whether ceiling effects are evident.

Acknowledgements

I would like to say a special thanks to my supervisors Ms. Julie Brinton and co-supervisor Dr. Carl Verschuur for all their guidance and support throughout this study. I would also like to extend my thanks to the staff in the South of England Cochlear Implant Centre (SOECIC), Southampton who took the time to participate in the video clips and discuss children from their case loads. Without your participation this study would not have been able to proceed. I would also like to thank all of the people who took the time to participant in this study. Many thanks also to my friends and family who continually supported me throughout the year.

Contents

List of Figures

List of Tables

Abbreviations

dB	decibels
CAP	Categories of Auditory Performance
CAP-II	Categories of Auditory Performance Two
CI	Cochlear Implant
GB	Gigabytes
HA	Hearing Aids
HI	Hearing Impairment
HZ	Hertz
K-S Test	Kolmogorov-Smirnov test
LiP	Listening Progress Profile
MAIS	Meaningful Auditory Integration Scale
NEAP	Nottingham Early Assessment Package
PASS	Profile of Actual Speech Skills
PALS	Profile of Actual Linguistic Skills
SIR	Speech Intelligibility Rating Scale
SNAP	Stories-Narratives Assessment Procedure
SLT	Speech and Language Therapist
SOECIC	South of England Cochlear Implant Centre
TOD	Teachers of the Deaf
TVA	Tait Video Analysis
UK	United Kingdom

1. Introduction

Over the past two decades there has been a rapid increase in the number of people fitted with cochlear implants (CI)s, especially young children in the United Kingdom (UK) (Archbold *et al.,* 1998). Children receiving CIs became fully established in Europe in the 1990's (Archbold & Robinson, 1997). The first child to receive a CI in the UK was in 1989 and the number of implanted children had increased to over 1,000 by 1999 and continues to increase (O'Neill *et al.,* 2002).

CIs are electronic implanted devices that substantially improve hearing in people with profound deafness (O'Neill *et al.* 2002). Profound deafness can result in deaf-mutism, a communication disorder which results from severe congenital hearing impairment (HI). As a direct result of this deafness, a child may experience absence of oral language development (Encyclopaedia Britannica, 2010). The prevalence of this condition is considerably high; approximately 1 per 2,000 babies born can be affected by this condition in the Western world (Mauk & Behrens, 1993; Parving, 1993; Davis & Parving, 1993; Stein, 1999). It results in extensive emotional, psychological and social harm (Schauwers *et al*., 2004a).

One method of currently managing profound deafness is the implementation of CIs which is considered a very valuable strategy at present, especially for those who receive little to no benefit from hearing aids (HA)s (Allen *et al*., 2001). However there are many factors that can have an effect on how successful they actually are. Some of the factors that apply, especially to children are the duration of sensory deprivation, age at implantation and the child's potential to develop generally (Huttunen *et al.,* 2009). This illustrates only a small number of factors that may have an effect and they need to be taken into consideration when determining whether or not CIs are the best method of implementation.

Evidence which shows that implanting young children is more beneficial for them has resulted in the growing trend in childhood implantation. The reason for this is because it is well established that the neural organisation and structure related to speech perception/production is affected by the length of auditory deprivation.

Although the extent of this is not fully understood and it is unknown whether there is potential for negative changes in the neural architecture to be reversed (Colletti, 2009). Due to this it is anticipated that the younger a child is when they receive a CI then it is more likely that they are going to truly benefit from it, especially in terms of developing language at a level that is appropriate for their age. Those under two years of age have revealed very fast improvements (Schauwers *et al.,* 2004b). The first year of a child's life proves crucial to their development. Children that are implanted at a young age usually show faster outcomes with use of their CI (Kishon-Rabin *et al.,* 2005).

Some of the significant impacts that CIs have on children with profound hearing are improved hearing, speech and language capabilities (Schauwers *et al.,* 2004a). However the evidence that supports the need for children to be implanted at a very early age is only slowly being obtained. This is because it is extremely difficult to correctly assess the auditory performance of very young children (Govaerts *et al.,* 2002).

As a result of the rapid increase in paediatric cochlear implantation there remains the need to provide sufficient methods in order to assess the benefits of this intervention. It has been recognised that there are gaps in the appropriate performance measures for linguistically developing children so parental interviews and rating scales are essential (Inscoe, 1999). Assessing the benefits of CIs in children is proving increasingly difficult since they are getting too young to correctly assess using clinical methods. Archbold and Robinson (1997) found that of the 504 children in Europe that were proposed to receive a CI during 1996, 54% of them were 2-5 years old and 12% were 0-2 years old revealing a trend towards implanting children at a younger age (Archbold & Robinson, 1997). This study was conducted over a decade ago and since then the number of children implanted under two years of age has grown.

The very young age of children that are now being fitted with CIs is not the only problem for assessing the benefits of this device. The development of a child can also be a factor, as it can differ between children. The amount of progress that they make after implantation can be very variable.

Even though there are methods of assessing a child's benefit with a CI there is a shortage of a sufficient means that can be applied successfully to children of all ages before and after they have been fitted (Archbold *et al*., 1998). Many methods are not suitable for assessing a child until they are over six months old (Nikolopoulos *et al*., 1999a).

Long-term studies are necessary to measure how a child is progressing with a CI as their development can extend over many years (Nikolopoulos *et al.,* 1999a). Many assessment methods are limited and are only suitable for measuring progress up until 5 years (Beadle *et al.,* 2005). However methods should be developed that can assess children for at least 10 years and beyond. Assessment methods that are long-term and suitable for assessing children up until 10 years are the Categories of Auditory Performance (CAP) and the Speech Intelligibility Rating scale (SIR) (Beadle *et al.,* 2005).

There are many aspects that affect the outcomes of paediatric implantation. Parental expectations, resources to support the child in acquiring auditory experiences, dynamics of social relations within the family, technological factors and quality of habilitation all have a significant impact. These are not the only factors, there are many more. Only some of the outcomes gained by having a CI can be assessed by clinical tests, such as behavioural tests.

Behavioural tests alone cannot truly illustrate the outcomes that the child experiences in their daily lives or their progress with CIs at home or in school (O'Neill *et al.,* 2004). For example a clinical behavioural test cannot emphasize whether a child has the ability to use the phone and discriminate speech with use of CIs. So therefore there is still need for assessing everyday communication, social relations, well-being and other constituents of quality of life of implanted children (Huttunen *et al.,* 2009).

Methods of doing this are interviews, open- or closed-format, semi-structured questionnaires or rating scales as they provide a broad picture of the social development of the child in different situations, at home or in school (Huttunen *et al.,* 2009).

These assessment methods which can be based on different professionals' views and parental perspectives are needed since there can be differences between the results obtained from clinical tests. Although there is the potential for parents to be biased, their views are valuable since they can add to clinical assessments carried out by professionals.

There are a lot of factors that need to be taken into account while considering a child for a CI. As the age at implantation is getting significantly lower, the procedure and assessments prior to and after implantation is proving extremely challenging. There is an increase in bilateral fittings and many factors regarding CIs are changing. One of these factors that are relatively new to CIs is insertion of a shorter electrode into the cochlea to restore high frequency HI whilst retaining residual hearing (NIDCD, 2008). Assessment methods need to be further developed and adapted in order to account for the new advancements of CIs and correctly assess the benefits that children are receiving.

1.1. Cochlear Implants

CI's are neural prosthetic devices that improve hearing by enabling partial restoration of cochlear function in people with a profoundly deaf sensorineural hearing loss (SNHL) (Schauwers *et al.,* 2004b). HI people who are eligible to receive a CI are those who receive very little to no benefit from conventional HAs (O'Neill, *et al.,* 2004). The implant is surgically implanted and the procedure involves inserting the electrode into the cochlea (Figure 1). When a sound stimulus is received, via the microphone it is then digitally analysed and the signal is converted by the processor into a coded signal (Schauwers *et al.,* 2004a). The coded signal then stimulates the spiral ganglion cells that innervate fibres of the auditory nerve, by-passing the damaged sensory structures of the inner ear.

Only the predetermined locations on the auditory nerve receive the coded signal (Archbold & O'Donoghue, 2009). This activation of electrodes provides a sensation of hearing but unfortunately does not restore hearing (NICE, 2009). CIs can provide aided threshold of approximately 25-30 dB across the speech frequencies. This gives the child access to the whole frequency range enabling them to hear and develop their speech. This level of hearing is very close to within normal limits (Archbold & O'Donoghue, 2009).

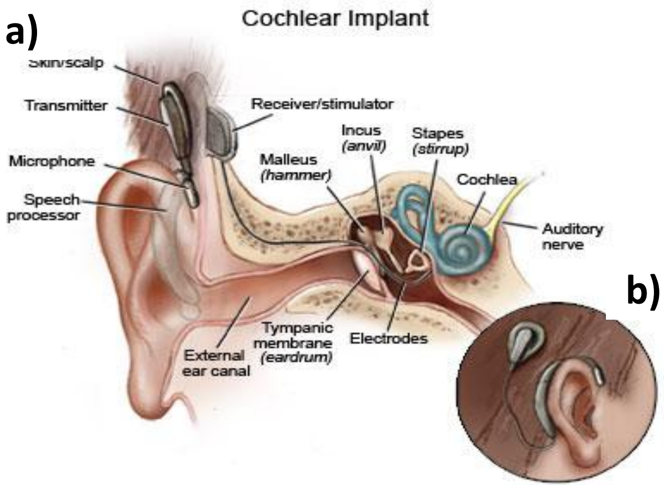

Figure 1: Diagram of a cochlear implant (CI) surgically implanted in the ear (a), the processor and transmitter which is external to the ear (b) (Adapted from The Nemours Foundation, 1995-2010).

Initially authorities doubted that CIs would be beneficial to young congenitally deaf children. They were sceptical as to whether young children would have the central processing abilities to enable them to hear speech through the implant and then produce intelligible spoken language. However research has revealed that the devices have the ability to enhance speech perception and speech production in these children (Nikolopoulos *et al.,* 1999a).

Now many children born with a profound congenital loss are implanted, as CIs have proven to provide the most significant change in the management of childhood hearing loss (Archbold & O'Donoghue, 2009). A study on 133 children found that those that were implanted and reaching the 6-year interval after implantation revealed a significant increase, 82% increase in the percentage of children that were able to understand speech without lip-reading (Nikolopoulos *et al.,* 1999b). This reveals the vast improvement of auditory perception after a CI is fitted.

1.1.1. Criteria for Selection for a Cochlear Implant

Currently in the UK CIs are only offered to children between 12 months to 17 years with profound SNHL. They only have the ability to hear sounds with an intensity ≥ 90 dB HL at frequencies of 2 and 4 kHz and gain no benefit from HAs (NICE, 2009), (Archbold & O'Donoghue, 2009). Adequate benefits from conventional HAs result in a child not being eligible for a CI. These benefits are classified as having listening, language and speech skills appropriate to their age, developmental stage and cognitive ability.

Children have to have at least a three month trial on a HA before they can be considered for a CI, as this is part of the assessment criteria unless there is a valid reason why it may be inappropriate (NICE, 2009). Now almost everyone who requires a CI is implanted with a multi-channel device and there is also an increase in bilateral fitting. There are currently 4 manufactures of CIs; the Nucleus device manufactured by Cochlear Ltd, Australia, the Advanced Bionics device made in the US, the MED-EL device, made in Austria and Neurelec manufactured in France (Archbold & O'Donoghue, 2009).

There has been much technological advancement in the speech processing of CI's over the past 20 years. In Beadle's *et al.,* (2005) study a large percentage of children could use the telephone, 60% after 10 years of CI use. This is a significant outcome and illustrates the importance of the device and also helps promote independence in adulthood (Beadle *et al.,* (2005). It also indicates that assessment measures also need to develop in order to assess the advanced developments of children with CI use.

1.1.2. Bilateral Cochlear Implants

Despite the increasing improvements of CIs, which leads to better language acquisition and reading skills, there are difficulties for children listening in background noise. This is especially a problem in class particularly for school age children.

Lack of the ability to take advantage of the better-ear effect could explain poor performance of HI people in noisy environments (Litovsky *et al.,* 2004b). It is well known that binaural hearing is important for speech-in-noise and localisation abilities. In an attempt to find a solution to this problem some patients have been fitted with bilateral CIs (two CIs). Evidence has shown that bilateral CIs restore the fundamental aspects of binaural hearing and the advantages that normal hearing-listeners experience (Kim *et al.,* 2009). Bilateral CIs are more commonly available for adults and less readily available for children (Litovsky *et al.,* 2004a). However, recently some children are now being fitted with bilateral CIs.

When CIs are fitted sequentially evidence has shown that speech intelligibility in noise is generally better when listening under bilateral conditions compared to either the first or second implanted ear alone. These improvements were found during the first 9 months of implantation. However more research into this and long-term measures for example the CAP scale are needed to assess these benefits (Litovsky *et al.,* 2004a). Some of the other benefits that can be achieved by having two CIs are more natural hearing, improved quality of life and also reduced effort and concentration when listening (Kim *et al.,* 2009).

Litovsky *et al.,* (2004a) found that the advantages i.e. speech intelligibility that children receive with bilateral CIs improve within the first 9 months of implantation. However Litovsky *et al.,* (2004a) also state in their study that follow-up measures post-implantation need to be carried out over a long period of time with more participants in order to determine the most important factors that contribute to the advantages of bilateral CIs in children.

They also found that using bilateral CIs may be very important for listening to speech in noise, negotiating complex acoustic environments but are not so important for listening to speech in quiet (Litovsky *et al.,* 2004a). In contrast to these findings Kim *et al.,* (2009) and Kuhn-Inacker *et al.,* (2004) found that there were significant improvements in speech perception in quiet as well as in noise with bilateral implantation when comparisons were made with unilateral CIs.

Children achieve much more additional value to their speech and language development as a result of bilateral CIs compared to adults. This could possibly be due to the fact that the central auditory system is highly plastic for a sensitive period of approximately 3.5 years in early childhood (Sharma *et al*., 2002c). So therefore if bilateral CIs are fitted within the first three and a half years of a child's life they will be more advantaged from bilateral implantation than adults or older children. Parents also found that their children showed advantages in their daily lives, improved articulation, and participated in conversations at ease (Kim *et al.,* 2009). As a result of this rating scales such as the CAP need to be further improved to account for these benefits.

When comparisons were made between the localisation abilities of children and adults with bilateral CIs, children appeared not to do as well as adults after 9 months of bilateral CI use. If children are studied over a longer period of time however, there may be evidence of improved localisation abilities with bilateral CIs (Litovsky *et al.,* 2004a). In another study Litovsky *et al*., (2004b) found that there was no significant difference in localisation abilities between monaural and binaural listening conditions in children. Even though these results were not significant, there was slight improvement noticed with the binaural condition (Litovsky *et al.,* (2004b). However only three children were assessed and this could be a factor that has demonstrated no significance in localisation ability in children. More research into this area needs to be carried out, assessing a larger number of children.

Much more research is needed into the benefits of bilateral CIs in children and generally bilateral fittings as a whole. In order to establish the extent of the advantages of speech perception in quiet and noise, long-term studies are needed as children may require a prolonged period of adjustment and learning (Kim *et al.,* 2009). It is not readily understood to which extent binaural cues are available to listeners and how effectively they are used which also emphasises the need for future research.

1.2. Inter-Rater Reliability

Inter-rater reliability is essential when developing subjective assessment methods and rating scales. Reliability between assessors is extremely important when conducting studies especially when comparisons are made between different participants. Reliability refers to the internal consistency and test-retest reliability of an assessment method. The assessment tool should yield reproducible measurements and generalisable inferences in a specific context in order to be considered reliable (Hyde, 2000). If the true state of the participant does not change then replicate measurements should be consistent on different occasions or with alternative versions of the assessment tool (Hyde, 2000).

High inter-rater reliability means that two or more assessors have categorised a large percentage of data the same way with only slight variations in the results. This means that the error variation due to those making the assessments is relatively small (Nicholas *et al*., 1999). If reliability estimates are low this can be an indication that there is a substantial amount of bias from the assessors that are making the judgements. This can result in poor representation of a child's performance and true abilities (Nicholas *et al*., 1999).

If assessors are all rating the same individual differently then it will not be possible to determine how well that individual is progressing with their CI especially if there are huge differences between the categories assigned i.e. when using the CAP rating scale. If two assessors assign a CAP score that differs by more than one category then it will be unclear as to how well the child is actually progressing. Acceptable reliability is crucial in order for an assessment measure to be useful (Hyde, 2000).

The items/categories must also have a balance between homogeneity and diversity. They need to form a scale. However it is insufficient to slightly change the wording of each category or by repeating items in the scale. If two categories are very similar then the responses will be highly correlated and less information will be gained from the assessment method (Hyde, 2000).

It can be very difficult for an assessor to make correct decisions regarding the performance of young children. Therefore it is important that the scale is sufficient in that it allows assessors to make accurate judgements about a child's hearing (Nicholas & Geers, 1997; Nicholas *et al*., 1994; Nicholas, 1994; Nicholas *et al.,* 1999). If the categories are very similar this can make it even more difficult for them to make accurate judgements as to how they feel the child is progressing.

A problem when determining inter-rater reliability estimates is that percent agreement is often used as the reported statistic. However it can be very difficult to determine whether these estimates are calculated on a point-by-point basis or some other basis (Nicholas *et al.,* 1999). Cordes, (1994) illustrates that percent agreement between inter-raters alone does not account for reliability. However more recently researchers are using Cohen's kappa statistic (Cohen, 1960) to correct chance of agreement (Nicholas *et al.,* 1999).

New assessment methods need to be evaluated very carefully before undergoing development (Hyde, 2000). Validity also needs to be taken into account as it is also important when developing assessment methods, as it reflects the ability of the method to yield relevant and useful information in different contexts. If a method is not valid then the reliability of the assessment method will be irrelevant (Hyde, 2000). Both reliability and validity are very important when developing assessments methods such as rating scales.

1.3. Floor and Ceiling Effects

New assessment methods must not show floor or ceiling effects the latter has become increasingly evident with the current CAP scale. Ceiling effects occur when the scores of all participants tested have almost reached the maximum level possible for that particular test or scale (Wikia Education, 2010). Participants are all achieving similar high scores which mean that comparisons cannot be made between them as they are all achieving the same. There is no room for measuring further improvements as the test or rating scale is limited.

For example in the CAP if all participants achieve the same score, a maximum score, of 7 and there are no further categories developed, then further improvements with CI use cannot be assessed.

Floor effects can also be a problem when designing assessment methods. Floor effects are the opposite of ceiling effects and refer to all participants achieving low scores on the rating scale and none of them are able to gain a higher score. This is because their performance is restricted by the limitations of the rating scale, and it is too difficult. Participants do not have the ability to do better and demonstrate progress (Answers.com, 2010).

1.4. Overview of Assessment Methods

At present there are very few assessment methods that are suitable for assessing the progress of children after they have been fitted with a CI, and even fewer measures are available for assessing very young children. The reason for this is because it is difficult to assess them and therefore behavioural assessments are very limited for this age group. (Nikolopoulos *et al.,* 1999b). Thus other subjective assessment methods are needed.

How a child appears to be performing and developing according to the clinical audiologist's assessments may not truly reflect how the child is doing at home. The teachers,' parents' and professionals' perspectives based on subjective assessment methods of a child's performance may be completely different from what the clinical tests reveal (Vidas *et al.,* 1990). These subjective methods are useful in that they provide information on the decision making process, functioning of the device, appropriateness of support and information that may provide help with identifying any additional learning difficulties (Nikolopoulos *et al.,* 2005).

However the methods that are used for reporting outcomes of children with CIs has been criticised due to serious weaknesses in the technique. Some of these weaknesses include selection bias, omitting low performers and not including individual data on all the participants, the protocol, poor study design and inadequate follow-up (Beadle *et al.,* 2005).

There are many different types of assessment methods that can be used to assess children although each of these is limited and should not be used alone. The Nottingham early assessment package (NEAP) (Figure 2) for monitoring progress in young children is an example of different methods that are used together in order to get a true indication of how a child is doing after they have been fitted with a CI (Nikolopoulos *et al.,* 2005).

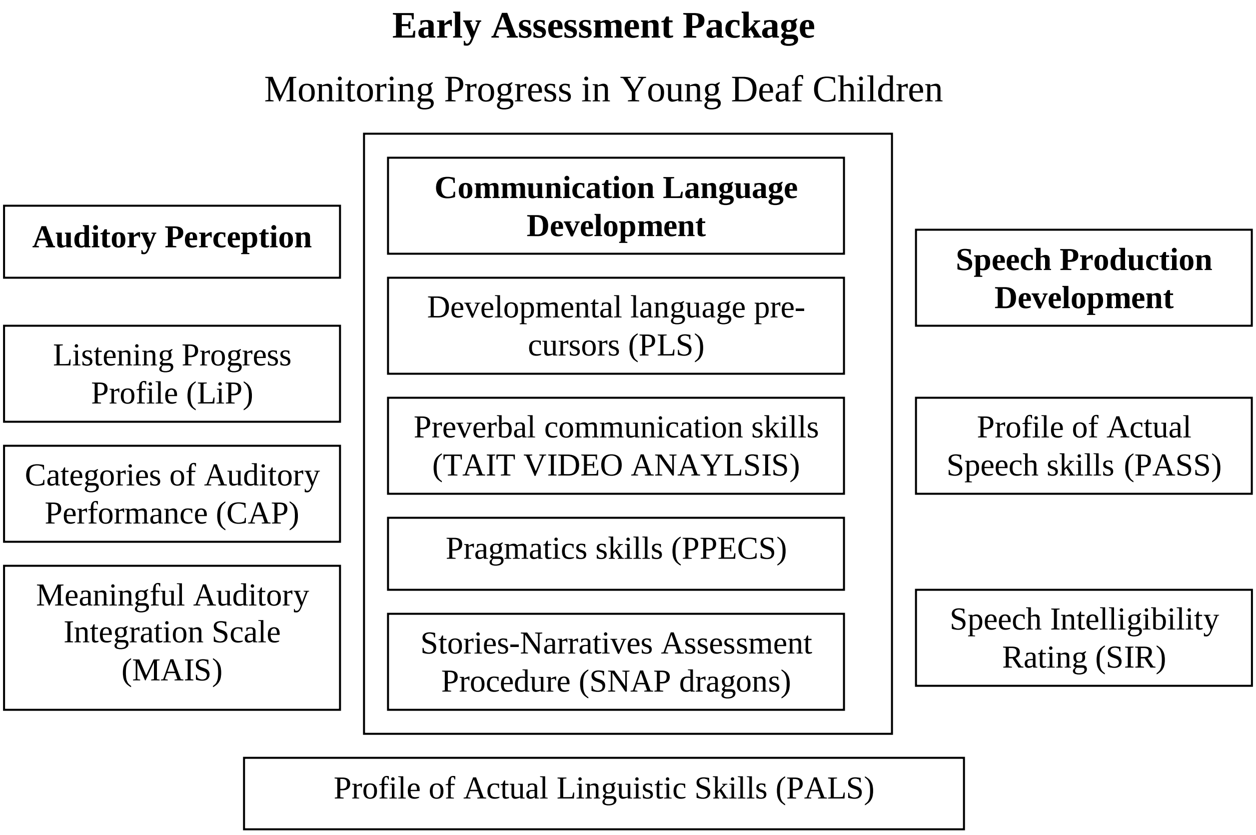

Figure 2: The assessment measures used in the Nottingham early assessment package (NEAP). This is a framework which can be used to assess the use of audition, language and communication of young children in real-life situations (Adapted from Nikolopoulos *et al.,* 2005).

Nikolopoulos *et al.,* (2005) conducted a study investigating the seven assessment measures that form part of the NEAP (Figure 2). These seven assessment measures include the listening progress profile (LiP), categories of auditory performance (CAP), tait video analysis (TVA), stories-narratives assessment procedure (SNAP Dragons), Profile of Actual Linguistic Skills (PALS), speech intelligibility rating (SIR) and the profile of actual speech skills (PASS).

The meaningful auditory integration scale (MAIS) has been modified to be included in the NEAP. Some of these assessment measures can be used regardless of mode or language and focus on communication itself (Nikolopoulos *et al.,* 2005). There are many other assessment tools which are not included in the NEAP framework.

The NEAP is a framework which is used to assess young children in their use of audition, language and communication skills in real-life situations. It provides a structured way of assessing the child at the pre-lexical stage before they are able to complete standardised tests when linguistic skills can be used (Nikolopoulos *et al.,* 2005).

The NEAP is effective in that it helps clinicians determine the appropriate intervention and management strategies that the child may need. It is a simple, reliable and time effective framework that can be used daily in clinics and can also help inform audiologists during the tuning process. The NEAP enables identification of additional problems that the child may have in their daily development. For example a child may appear to have good auditory and communication skills but may not be developing speech. (Nikolopoulos *et al.,* 2005).

Some assessment methods are only suitable for monitoring the child's progress over a short period of time (TVA, LiP, PASS) whereas others can be used more long-term (CAP, SIR and SNAP) (Nikolopoulos *et al.,* 2005). In order for studies to be effective over a long period of time the assessment methods need to have a wide range of outcomes to avoid floor and ceiling effects which become apparent quite quickly for some of the short-term measurements. Children's performance and speech intelligibility can continue to develop after 5 years of CI use. Although very few studies investigate performance beyond 5 years (Beadle *et al.,* 2005).

Monitoring scales/assessments must be scaled in such a way that spreads milestones relatively evenly across the range (Archbold *et al.,* 1998). Evidence has shown that more long-term follow-up of children fitted with a CI is needed to give a true indication of how they are developing after receiving an implant. 1 in 2 children appear to be developing and improving their speech after the 5 year interval post implantation (Beadle *et al.,* 2005).

Due to this evidence CAP should be used regularly to supplement other tests (Nikolopoulos *et al.,* 1999a). Further reasons for global long-term follow-up of children with CIs such as the CAP are to illustrate the cost-effectiveness of CIs, to describe the social benefits, enhanced educational attainment, employability, less dependence on psychiatric services and greater participation in social situations and peer-fairness among those who receive implants (Beadle *et al.,* 2005).

Although there are many different assessment methods for assessing children with CIs, only a few have been mentioned in this document, those that form part of the NEAP. However the main focus of this study will be on the CAP rating scale.

1.4.1. Categories of Auditory Performance (CAP)

CAP assesses the child's functioning in everyday situations, at home and in school. It covers a vast range of auditory performance and also takes into consideration the different developmental rates of children (Nikolopoulos *et al.,* 1999b). It is a global assessment method and has been used in many countries e.g. Europe and America. It covers a range of abilities which includes awareness of environmental sounds through to using the telephone and these are organised as a hierarchy of skills (Table 1) (Nikolopoulos *et al.,* 1999a).

Table 1: Categories of auditory performance (CAP) (Adapted from Nikolopoulos *et al.,* 2005).

Category	Criteria
7	Use of telephone with known listener.
6	Understanding of conversation without lip-reading.
5	Understanding of common phrases without lip-reading.
4	Discrimination of some speech sounds without lip-reading.
3	Identification of environmental sounds.
2	Response to speech sounds (e.g. ‘‘go’’).
1	Awareness of environmental sounds.
0	No awareness of environmental sounds.

The assessment involves observing the child using standard criteria and the categories have increasing difficulty which is useful in all stages of assessing the child's hearing skills as illustrated in table 1. It is easy to use and the inter-rater reliability has been validated as there was concern that different assessors may rate the same child differently limiting its usefulness (Archbold *et al.,* 1998).

CAP enables the assessors to gain insight into the functional use of audition in everyday life when CIs are fitted. The benefits achieved by having a CI over a period of time can be shown by the CAP score (Figure 3), which can measure development of speech recognition ability and functional hearing (Huttunen *et al.,* 2009). It is designed to give a readily assessable measure to non-specialists including parents and ordinary people who have no experience of assessments of HI children (Archbold *et al.,* 1995; Archbold *et al.,* 1998; Nikolopoulos *et al*., 1999b).

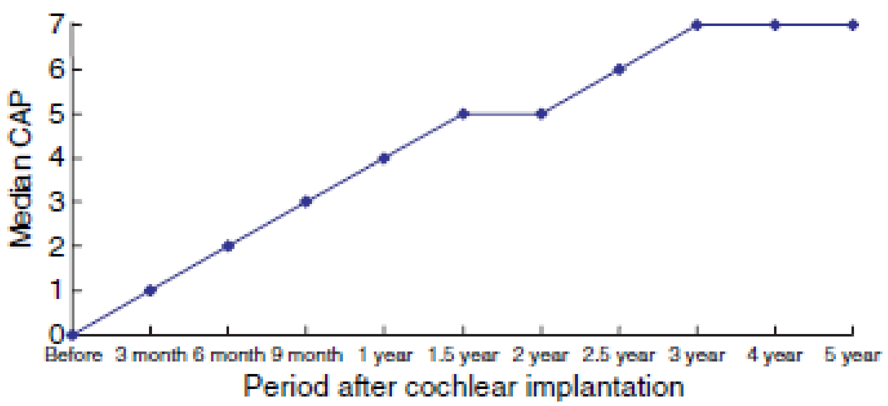

Figure 3: The development of median categorical auditory performance categories (CAP) versus time after cochlear implantation (Permission received from Wu et al., 2007). This figure shows that over time there is an increase in the CAP score. At three years of cochlear implant use the graph starts to level off as it has reached the maximum category.

It is an ordinal, nonlinear scale of auditory receptive abilities (Archbold *et al*., 1995, Archbold *et al.,* 1998; Nikolopoulos *et al,* 1999a). This means that it cannot be assumed that categories 1 to 3 are equivalent to a change in performance as categories 2 to 4. It is also unknown how much superior one category level is from another (O'Neill *et al,* 2002). Even though category 7 is rated better than category 6 it is not apparent how much better this level may be. It is important to take these factors into consideration whilst using CAP.

CAP is a long-term assessment method and has been proven very useful for assessing children of all ages both before and after implantation. It continues to assess auditory performance of children as they progress through to adulthood (Beadle *et al.,* 2005). Wu *et al.,* (2007) found that there are significant differences in the CAP score before implantation and one year after (figure 3). There are also significant changes 3 years after implantation (figure 3) (Wu *et al.,* 2007). Implantation consists of three key elements assessment, surgery and rehabilitation all these aspects can have an impact on the outcomes of CAP gain (O'Neill *et al.,* 2002).

Archbold *et al.,* (1998) found that the repeatability of CAP was demonstrated across a range of variables from before implantation until 5 years later. This is a crucial factor of the CAP as more long-term studies are needed to assess children's development after implantation. It can be used to follow-up a large group of children over a long time period (Archbold *et al.,* 1998).

A study carried out by Archbold *et al.,* (1998) revealed that there is good correlation between the inter-rater reliability. They found in their study that teachers who see the child daily compared to teachers in the implant centre, who did not see the children as often, rated the child similarly. This reveals that the CAP is reliable and gives a repeatable outcome regardless of where it was completed, either in a clinical-based situation or from observations in daily life. Although there were a few discrepancies in the score, they were very small and did not have a significant effect (Archbold *et al.,* 1998).

This method is useful for assessing children who are too young in order for behavioural audiological assessment to be carried out, though it cannot completely replace clinical assessments. Even though it is useful for assessing young children it can also complement formal audiological assessments that are carried out on older children (Nikolopoulos *et al.,* 2005).

It also should not be used in isolation as it does not provide a true indication of all the outcomes that may be achieved with a CI of its own merit. It should be used with other assessment methods such as those mentioned in the NEAP. Huttunen *et al.,* (2009) found that general functioning of a CI was not found to be associated with speech recognition score or with the CAP score as measured in the clinic. Instead the higher CAP scores assigned by parents were found to be associated with how the child was progressing in school/preschool as opposed to a clinical setting (Huttunen *et al.,* 2009).

Although CAP is a very useful measure there are various factors that can affect it as O'Neill *et al., (*2002) illustrated in their study. The older the child is at implantation then the lower the CAP gain was. Age at implantation and CAP are negatively correlated this also applies to the age of implantation and the number of medical consultations. The more recently a child is implanted then the less likely it will be that the child will receive a high CAP gain (O'Neill *et al.,* 2002).

In order to overcome these problems O'Neill *et al.,* (2002) suggested that early detection is necessary for HI children in order for early implantation. As a result of this there will be a higher chance that the child will be able to achieve higher CAP gain (O'Neill *et al.,* 2002).

Another study carried out that investigated the use of CAP showed that children are more likely to be placed in higher categories if they possessed attributes in four key domains (Stacey *et al.,* 2003).

These include:

1) (A child) "with a better average hearing level, that is older at the age of onset of HI, fewer additional disabilities aside from the HI;"
2) (Family) "higher socio-economic status, parents having milder degrees of hearing loss;"
3) (School) "teachers only using spoken language as opposed to combinations of speaking and signing" and
4) (Implantation) "possessing an implant" (Stacey *et al*., 2003).

Stacey *et al.,* (2003) found from their study, using the CAP rating scale that the influences on auditory performance among HI children are multi-factorial. Provided that all variables listed above are kept constant or are controlled then CIs have a significant positive effect on the child's hearing and performance.

In Huttunen's *et al.,* (2009) study they found that Finnish children in their study who used a combination of sign language and speech had significantly better CAP scores than those that just used sign language. Another study found that a child who used oral communication before implantation also performed better with their CI than those that used sign language as a means of communication (Taitelbaum-Sweed *et al.,* 2005). These two studies illustrated that mode of communication is a factor contributing to CI success. This is probably because children using sign language are depending on that more and are not relying so much on their speech. They are not encouraged to develop their listening skills and speaking skills to the same extent as other children that are using both speech and sign language.

Evidence in a study of 21 prelingually HI children showed that the median Categorical Auditory Performance after one year of implantation was a score of 4. After five years post implantation 81% of children had a CAP score of 7 (figure 3) (Wu *et al.,* 2007). However there are no higher categories so it is unknown whether the HI children could improve and have an even higher CAP score. This is indicative of ceiling effects that can occur because the rating scale is limited.

However Beadle *et al*., (2005) carried out a longitudinal study that investigated 30 implanted children before implantation, at 5 and 10 years after implantation. They found that after 5 years of implant use 93% of children had a CAP score of 5 and above, although only 31% of these children had the highest CAP score, category 7 (Beadle *et al.,* 2005). Some participants also appeared to plateau before they reached the highest CAP score, 5 years post-implantation in Beadle's *et al.,* (2005) study. This is considerably lower than what Wu *et al.,* (2007) found.

However 10 years after implantation the children in Beadle's, *et al.,* (2005) study continued to progress and their CAP scores increased. 60% reached the highest CAP score, category 7 and were able to use the telephone with a familiar speaker. This cannot be compared to Wu's *et al.,* (2005) study as they did not continue follow-up beyond 5 years post implantation.

These differences could be due to differences in language as Wu *et al.,* (2007) found that Mandarin-speaking participants seemed to develop quicker than English-speaking children and their CAP score increased much more rapidly. Whereas in Beadle's *et al.,* (2005) study it was carried out on English-speaking children. Another reason for the differences could be due to the age at which the children were implanted. In Beadle's *et al.,* (2005) study the children were implanted at a mean age of 5.2 years of age whereas the children in Wu's *et al.,* (2007) study were implanted before the age of 3. There is considerable difference in the average age that children are implanted in these two studies. This can have significant impact on the results and negatively affect how quickly a child achieves a high CAP score. One of the main factors that affect the CAP and a child's development and progress with a CI is the age that they are implanted.

CIs are continually being developed. Advanced features and technology is continually added to improve CIs for HI people. These improvements lead to increased benefits and positives outcomes of CIs. Children that are implanted are younger and there is also an increase in bilateral fittings which also leads to further improvements. As a result, outcome measures, as assessed by the CAP score are becoming better than what the current CAP rating scale can assess. The current CAP will soon show ceiling effects, there is evidence to show this is occurring already. It will not be sufficient to measure all possible outcomes that may be achieved with the new advanced CIs and bilateral fittings.

Therefore the original authors of the CAP have proposed that two new categories should be added to the current CAP scale to increase the number of outcomes that can be assessed using this rating scale (Table 2). This new CAP scale will be called categories of auditory performance two (CAP)-II (Table 2).

Table 2: Categories of auditory performance-II (CAP)-II. This table shows the two additional categories that are added to the CAP rating scale (categories 8 and 9 as shown in *italics*) (Adapted from Nikolopoulos *et al.,* 2005).

Category	Criteria
9	*Use of phone with unknown speaker in unpredictable context.*
8	*Follows group conversation in a reverberant room or where there is some interfering noise, such as a classroom or restaurant.*
7	Use of telephone with known listener.
6	Understanding of conversation without lip-reading.
5	Understanding of common phrases without lip-reading.
4	Discrimination of some speech sounds without lip-reading.
3	Identification of environmental sounds.
2	Response to speech sounds (e.g. ‘‘go’’).
1	Awareness of environmental sounds.
0	No awareness of environmental sounds.

1.5. Age at implantation

There is much debate as to how early a child should be fitted with a CI if they are found to have a profound HI. Evidence has shown that a profound hearing loss can result in delays in the onset of babbling in HI infants for at least 5-19 months (Oller & Eilers 1988; Oller *et al*., 1985) and in turn delay their speech development. Young children tend to lose their sensitivity for non-native speech contrasts. This language specific discrimination capacity has important consequences for a child’s perceptual or auditory functioning and for their speech and language development. Therefore children implanted as early as possible within the first 12 months of life would be able to develop language and speech similar to that of their hearing peers (Schauwers *et al.,* 2004a; Kishon-Rabin *et al.,* 2005).

Schauwers *et al.,* (2004a) found in their study that 8 of the 10 children, implanted between 6-18 months achieved a relatively high CAP score (5 or 6) one year after implantation.

O'Neill *et al.,* (2002) found that there is a negative correlation between CAP gain and age at implantation. Schauwers *et al.,* (2004b) and Govaerts *et al.,* (2002) also found that audiological outcomes decrease with age of implantation. In order to overcome this they have suggested that the earlier a child is implanted then the more likely they will progress and as a result their CAP scores will increase (O'Neill *et al.,* 2002).

Most children however are only fitted with a CI after they have had a minimum trial period of three months with a conventional HA which has been found to be unsuccessful. The standard guidelines at present only offer CI to children over 12 months of age (NICE, 2009). However recent research has revealed that a child could benefit more, in terms of spoken language if they are fitted with a CI younger than 12 months (Colletti, 2009).

Age has been considered a significant factor in predicting the outcomes in paediatric CIs in many studies although there is only a little statistical data in the literature to support this claim (Nikolopoulos *et al*., 1999a). Apuzzo and Yoshinaga-Itano (1995) found that children identified as having a hearing loss and aided in the first two months of life develop significantly better language than those aided between 3-12 months of age. Colletti *et al.,* (2005) also found that the onset of babbling was very early in infants within 1 to 3 months after implantation than those fitted less than 1 year of age. However this study size was very small and these results were shown in all 10 infants included in the study.

Colletti (2009) also states that levels of spoken language competence, breadth of vocabulary and the complexity of sentences appear to be directly affected by the age a child is fitted with a CI. The performance of children implanted very early was very similar to that of their normally hearing peers (Colletti, 2009).

The group of children implanted between 4-11 months of age revealed the most rapid increase in CAP scores compared to the two other groups of children that were implanted later, between 12-23 months (group 2) and between 24-36 months of age (group 3). Group 1 reached the highest CAP score possible; scoring 7 at their 24 month follow-up.

The other two groups reached this level later at 36 months (Group 2) and 42 months (Group 3) after surgery (Colletti, 2009). Schauwers, *et al.* (2004b) also found that children implanted in the first year of life reached a normal CAP score as early as 3 months after implantation, whereas children implanted in their second year of life needed up until 12 months to achieve a CAP level appropriate to their age.

If age is found to be a major predictor of long-term functional outcome then it would have major implications for when a child should be implanted (Nikolopoulos *et al*., 1999a). Nikolopoulos *et al*. (1999a) have carried out research into the importance of age for paediatric implantation to investigate the influence in speech perception and intelligibility. O'Neill *et al.,* (2002) stated that it can take a long time for improvement in outcomes to become apparent.

Four different assessment methods including the CAP rating scale were used in Nikolopoulos *et al.,* (1999a) study to investigate 126 congenitally and prelingually deaf children. The reason why different assessment measures were used is because they all measure different things and are limited on their own as to what they can tell us about how a child is progressing. So these different measures were used in combination to give a more accurate picture of the results with age at implantation (Nikolopoulos *et al*., 1999a).

The onset of children's deafness in this study was under three years of age and they were implanted under the age of seven. The causes of deafness varied between individuals, some causes were genetic, 59% whereas others were caused by other factors such as meningitis. They were assessed for up until 4 years post implantation (Nikolopoulos *et al*., 1999a). The results revealed that there is a positive correlation between the outcomes and age at implantation (Nikolopoulos *et al*., 1999a).

However the results showed that children implanted at three and four years of age had negative correlations for all three of the assessment methods used, especially for CAP (Nikolopoulos *et al*., 1999a).

Children implanted between 2-4 years of age can still be at risk of significant and irreversible delays in various aspects of their development (Schauwers *et al.,* 2004b). As they are not getting sufficient aid and are not receiving available sounds that are essential for speech development in their prelexical stage which usually averages around development at approximately 2 years of age (Miyamoto *et al.,* 1997). In support of this Nicholas and Geers (2007) states that the likelihood that a child achieving normal language in preschool period decreases the older they are implanted. Children that are not fitted until three years of age may experience extreme difficulty catching up with their hearing peers (Nicholas & Geers, 2007).

However Govaerts *et al.* (2002) state that children can eventually yield good auditory outcomes after implantation between the ages of 2-4 years of age although it may take 3-4 years for this to happen. Govaerts *et al.* (2002) also found that children implanted after 4 years of age only have a small chance (20-30% chance) of achieving normal CAP scores. This illustrates the importance of age at implantation and the impacts it may have on a HI child's development.

Overall the results reveal that the earlier a child is implanted the more they are going to benefit from their CI and the more likely it will be that they will be able to attend mainstream school (Schauwers *et al.* 2004b). However this poses real problems methodologically especially in terms of assessing the auditory performance of very young children reliably. This in turn highlights the importance of indirect measures of assessing children for example rating scales such as CAP (Archbold *et al.,* 1995).

The problem with Nikolopoulos' *et al*., (1999a) study is that there could be variations in the numbers of children implanted at each particular age. If there were considerably more children implanted at two years of age compared to other age groups i.e. six years of age then the results may not truly reflect the average scores that the children are achieving.
Also it is unknown whether all participants have developed normally despite their hearing disability. If some children have other developmental disorders then this could have an effect on their ability to perform well in the assessments.

There are many limitations of the studies that have been conducted on age at implantation. Some of these studies include short follow-up time and relatively small series. Govaert *et al.* (2002) only followed up their participants for two years after implantation. They stated that the children implanted after four years of age did not reach the highest CAP level, category 7 within the first two years after implantation (Govaert *et al.,* 2002). It is unknown whether these children eventually reach the highest CAP score as follow-up has not be conducted beyond two years. This has resulted in conflicting conclusions.

The results of Nikolopoulos' *et al*., (1999a) study reveals that long-term follow-up is needed in order to determine the true effects of age and account for a child's hearing development, as it can take many years to complete development (Nikolopoulos *et al*., 1999a). Stacey *et al.,* (2003) stated that the younger children are implanted and the longer they use their CI the greater the influence will be and their CAP scores will be higher.

Despite the limitations of studies conducted on age at implantation Nikolopoulos *et al*, (1999a), in conclusion to their study found that age of implantation was a strong predictor of outcome with regard to speech perception and ability in prelingually deaf children who were implanted under seven years of age (Nikolopoulos *et al*., 1999a).

A larger group of children diagnosed and tested longitudinally with HAs and CIs may highlight the importance of early identification and rehabilitation of children with a hearing loss (Kishon-Rabin *et al.,* 2005). More research into this area is needed in order to ascertain the importance of early implantation.

1.6. Conclusion

Many factors need to be taken into account whilst fitting CIs in children and assessing them. Only a few of these factors have been discussed here in detail and many other aspects need to be reviewed. One of these factors includes the age that they are implanted. Although a lot of researchers' state that age of implantation is a major factor, there is only a small amount of statistical data to reinforce this view. Nikolopoulos's *et al*., (1999a) study on age of implantation illustrated the positive effects of implanting very young children using 4 different assessment measures. However more research is needed in this area.

Currently children are not implanted in the UK until after a 3 month unsuccessful trial with HAs. They are usually not implanted until they are 1 year old or older. However if the current guidelines change and it is recommended that children should receive CIs under 1 year of age then assessment methods will need to be further developed, in order to account for the benefits they are gaining more rapidly and also for any additional benefits they may receive i.e. using the telephone with an unknown speaker. Evidence has shown that children are now able to hear things and progress with CIs in ways that professionals would never have begun to imagine a few years ago.

Research has also shown that there is lack of follow-up after CIs have been fitted. Most studies do not review children beyond five years of CI use. However there are a few studies that have followed up implanted children for 10 years. These studies have shown that children still have the potential to progress with their implant beyond 5 years. Ideally children should be followed up throughout their school lives to ensure that they are doing as well as possible and gaining maximum benefit. Sufficient assessment methods are needed in order for this to be possible.

Long-term follow-up of children fitted with CIs is essential in order to monitor, continual use, device failure and the outcomes of CIs (Beadle *et al.,* 2005). Programs need sufficient resources in order to assess these areas over a long period of time.

Few of the assessment methods available at present are suitable for this and many are only short-term measures that show ceiling effects after a couple of years post-implantation. However the CAP rating scale proves a very valid and important scale that is sufficient for this purpose although recently clinical evidence has shown that it needs adapting. As it has become limited in its ability to accurately assess the benefits that young children are receiving with their CIs (Beadle et al., 2005).

There has been increasing improvements in CI technology such as advances in the speech processing strategies that are used in CIs, although very little is known yet about how much benefit these provide to a CI user. Measures of speech perception, intelligibility and language acquisition, and other factors have been studied extensively though very little is known about a child's ability to function in noisy environments with CIs indicating that more research in this area is also required (Litovsky, *et al*., 2004b). As a result of these increasing improvements the benefits that children are receiving with CIs is beyond the scope of the original CAP rating scale.

One of the most recent advances in CIs for children is an increase in bilateral fittings. The extent to which binaural cues are available to listeners and used effectively is not very well understood indicating that the knowledge in this area is very limited and much more investigation is needed. Future work needs to address the extent to which synchronisation and enhanced binaural hearing are functionally useful to those using bilateral CIs (Litovsky *et al.,* 2004b).

Young children are very difficult to assess using clinical tests alone so other subjective methods are required. These subjective measures essentially include rating scales such as the CAP. It is vital that rating scales also improve along with recent advances in CIs. Additional categories should be added if necessary and were feasible in order to correctly assess the additional progress that children may be making with their CI. They need to be sufficient enough to account for all possible positive outcomes.

Due to these factors it is proposed that the CAP should be improved as ceiling effects have started to become apparent in a few studies and this limits the use of the scale, especially with young children and those fitted with bilateral CIs.

Therefore additional categories have been added to the CAP rating scale and they will be investigated to see if they are sufficient for assessing continuous progress of children after CIs have been fitted. Inter-rater reliability is an important factor when developing new rating scales so this needs to investigated and taken into account when making adjustments to the current rating scale.

1.7. Research question

Based on all the evidence discussed and the limitations that are becoming more apparent with rating scales, one of the original authors of CAP, has proposed the inclusion of two additional categories for the CAP-II rating scale (Table 2). This new scale will be referred to as CAP-II. These new categories are more difficult than the highest categories currently in the CAP rating scale. Category 9 is the considered the hardest category and it may be difficult for some children to achieve a CAP-II score at this level.

The two additional categories are as follows:

1. Category 9: '*Use of phone with unknown speaker in unpredictable context',*
2. Category 8: '*Follows group conversation in a reverberant room or where there is some interfering noise, such as a classroom or restaurant'.*

1.7.1. Is the inter-rater reliability of the (CAP)-II rating scale high?

The study will aim to establish inter-rater reliability of a large number of participants. This is justified in that this study wishes to ascertain whether there is high inter-rater reliability in the CAP-II rating scale, especially between the last three categories (category 7 to category 9) (Table 2).

1.7.2. Is the (CAP)-II rating scale a suitable assessment method for children with cochlear implants avoiding ceiling effects?

The study will also investigate whether there are ceiling effects found with the CAP-II rating scale when young children with CIs are assessed. It will take into account whether or not participants rate all children with CIs at the highest CAP level, category 9 or whether the children are all rated differently.

1.7.3. Are the two new proposed categories suitable to cover the increasing improvements of CIs?

Observing the responses of participants and taking into account their comments given throughout the study, it will be considered whether the two additional categories in the CAP-II rating scale are suitable for assessing young children with CIs or whether they should be developed further

2. Methodology

2.1. Aims

The aims of this research were to set up a viable procedure to assess the new version of the CAP with additional categories added to it, in order to evaluate the inter-rater reliability of the CAP-II rating scale for future assessments of children. In addition, another aim was to determine whether the two additional categories are sufficient for assessing young children with CIs and those that are bilaterally fitted whilst avoiding ceiling effects.

2.2. Hypothesis

2.2.1. Hypothesis 1

Based on previous research (Archbold *et al*., 1998) it is predicted that inter-rater reliability will of the CAP-II rating scale will be high.

2.2.2. Hypothesis 2

It is proposed that the CAP-II rating scale with the additional categories will not show ceiling effects.

2.3. Research design

2.3.1. Sample size

The number of participants included in the study was based on professional guidance and other previous research (Archbold *et al*. (1995) and Archbold *et al*., (1998). This study initially aimed to recruit 30 healthy participants with good hearing, preferably all with English as their first language. However it proved difficult to find enough participants with this requirement. Three extra participants volunteered and a total of 33 participants were included in the study. Participants were recruited from amongst the staff, student population and their friends and family at the University of Southampton.

2.3.2. Inclusion criteria

The aims of the set criteria was to ensure that participants included in the study understood fully what was required of them and to represent as accurately as possible reliability between participants rating the videos. This was considered necessary, as participants with different backgrounds and language differences may misinterpret what each category means in the rating scale. They may also misinterpret what is said in the video clips, therefore rating each child differently to participants who have English as their first language.

Participants were selected according to the following selection criteria:

- Healthy individuals with no hearing or other health problems.
- An excellent understanding of the English language. This was to ensure that they have the best possible chance of understanding the CAP-II rating scale and what is said in the videos.
- Age 18 years+ (Parental consent would not be required).
- Both genders.
- Any social background.
- No extensive experience of CIs or CI users.

2.3.3. Equipment

Six professionals working in the South of England Cochlear Implant Centre (SOECIC) were invited to talk about non-identifiable children from their case loads. This was done as opposed to using recordings of real children or their parents discussing their children due to ethical issues. It would take a long time to take to get ethical approval to use recordings of children or parents discussing their children. So recordings of professionals were used as the CAP-II rating scale is used by both professionals and parents.

Video recordings were made based on a guided interview of these 6 professionals (3 Speech and Language Therapists (SLT) and 3 Teachers of the Deaf (TOD)) responding to the CAP-II rating scale answering questions about children known to them but not identifiable (Appendix 1). The video recordings were made by a colleague in the SOECIC, a fellow Speech Therapist who guided the interviews and took each of the professionals through the CAP-II rating scale whilst they discussed each individual child in an anonymous way. The video clips were made using a Sony DCR-SR55 Handycam camcorder.

2.3.3.1. Specifications of the camcorder (Price runner, (1999-2010)):

- Memory size of the camcorder was 40 GB.
- Screen resolution was 123000 pixels.
- Digital zoom was 50 x.
- Optical zoom was 25 x.
- Still picture resolution was 1152 x 864 pixels.

2.3.3.2. Specifications of laptop

A laptop was used in order for participants to view the video clips on whilst taking part in the study.

The specifications were as follows:

- Hi-Grade laptop, the model number was NOTINO-W59001.
- Notebook computer, model W76T.
- Product code was W760T.
- Microsoft Windows 7, Windows Media Player 12 was the programme used to view the video clips (Windows, 2010).

2.3.3.3. Specifications of Speakers

Additional speakers (Creative SBS 50) were also used to produce good sound quality and ensure clarity of the interviewer and interviewees in the video clips.

The specifications were as follows (Cnet reviews, 2010):

- Audio output details/speaker output power was 5.5 Watt.
- Impedance was 7.5 Ohm.
- Output power/total was 11.0 Watt.
- Speaker type was Tweeter, right/left channel.
- Sound output mode was stereo.
- Power was 50/60 Hz.

2.3.4. Video Clips

- **Introductory Video** This video was prepared to enable participants to gain some ‘training’ of what was required of them. They were required to watch this video before beginning the study. This gave participants the opportunity to adjust the volume and screen to a comfortable level and it enabled them to get a feel of what they were required to do. The script of this video is included in appendix 2.

- **Video clip 1 (112** (MMH)**)** was based on a boy of 10 years 9 months of age who has had one CI for several years in his right ear which was fitted on the 23rd January 2007. He has therefore had his CI for 3 years and wears it all the time. The type of CI that he has fitted is one of the Cochlear devices. The aetiology of this child’s HI is unknown. The interviewee that was discussing this child was a TOD.

- **Video clip 2 (118** (KH)**)** was based on a 15 year old girl who is currently doing her GCSEs and has bilateral CIs. The aetiology of her HI is unknown. Her first CI was fitted in her right ear in May 1998, 12 years ago. She was fitted her second CI in December 2009, 8 months ago but relies on the first CI more and she wears it all the time. The type of CIs that she has fitted is the Cochlear devices. The interviewee discussing this case was a SLT.

- **Video clip 3 (115** (TT)**)** was based on an 18 year old boy who is currently doing his AS levels. He has 2 CIs one was fitted on the 26th September 2009; almost a year ago and the second CI was fitted about 6 months ago. The type of CIs he has are the MED-EL devices. Before he was fitted with CIs he was a HA user for a long time as he has had a deteriorating hearing loss from 5 years of age but had a sudden drop in his hearing from 12 years of age with no known cause. He was discussed in this video by a TOD.

- **Video clip 4 (111** (SCP)**)** was based on a 9 year 2 month old girl who has a second CI but uses her first CI all the time. The first CI was fitted in her right ear on the 26th September 2003, 3 years ago and the second one was fitted on the 30th April 2010, 4 months ago. The type of CIs that she has fitted are the Cochlear devices. The aetiology of her HI is Wardenburgs Syndrome. She was discussed by a SLT.

- **Video clip 5 (117** (DM)**)** was based on a 5 year 1 month old girl who had bilateral CIs at 1 year 7 months of age. The aetiology was caused by meningitis at 1 year 2 months of age. The CI in the left ear was removed in May 2009 following an infection. The left ear was re-implanted in November 2009, 8 months ago and it was retuned in January 2010, 7 months ago. The type of CIs that she has fitted are the Cochlear devices. She was discussed in this video by a TOD.

- **Video clip 6 (113** (AES)**)** was based on a 9 year 1 month old girl who was fitted with a CI in her right ear on the 2nd January 2004, 6 years ago and she also had a HA at this time. She was fitted with a CI in her left ear on 13th January 2010, 7 months ago. The type of CIs that she has fitted are the Cochlear devices. The aetiology of her HI is unknown although it is suspected to be genetic as her sister also has a HI. She was discussed in this video by a SLT.

2.3.5. Recording of Responses

Before the study was conducted a set of instructions were prepared that explained clearly what was required of each participant (Appendix 3). Participant record sheets were also prepared that each participant was required to fill in whilst doing the study (Appendix 4).

The participant record sheets included the CAP-II rating scale and a section for them to fill in any additional comments they may have about each video clip or the CAP-II rating scale (Appendix 4). An investigator's record sheet was also prepared (Appendix 5). One of these was used for each participant. Any repetitions that participants required whilst doing the experiment, any interruptions or anything that the participant needed clarified were noted down by the investigator on the investigator's record sheet (Appendix 5). All data that was collected was put into tables and graphs.

2.3.6. Procedure

Once safety and ethics approval was gained from the human experimentation committee and the safety and ethics committee at the human sciences group, Institute of Sound and Vibration Research, University of Southampton (Appendix 6) and the equipment had been safety checked by the safety committee, the study was conducted and data collection commenced.

2.3.7. Data Collection

Participants were seated comfortably, facing a computer screen in a quiet room where there would be minimal background noise and interruptions. Each Participant was required to watch an introductory video first whereby they were instructed to adjust their seat and screen to a position most suitable for them (Appendix 1).

Each participant was then given a set of instructions which they were required to read before beginning the experiment (Appendix 3). All participants were given the same set of instructions in order to avoid any bias in the experiment. They were also given 6 copies of the participant record sheet which included the CAP-II rating scale on it, 1 sheet was required for each video clip (Appendix 4). They were required to read and familiarise themselves with the CAP-II rating scale before beginning the study and rating each video (Table 2).

Participants were not told very much about what the experiment was or the uses of the rating scale as the investigator did not want to influence the responses of participants in any way. Only if participants needed anything clarified or had any questions then these were answered before they started to watch the 6 video clips to ensure that each participant understood exactly what was required of them.

Once participants were comfortable and ready to begin they were required to watch all 6 video clips (each approximately 3-5 minutes long, all videos including the introductory video took approximately 20 minutes in total to watch) which were in a randomised order. The video clips were randomised using a Latin square design to prevent bias (Williams 1949). Each individual was given their own unique video order and they were required to watch the videos in the specific order that they were given (Appendix 7). All videos were viewed on a Hi-Grade laptop, using Windows Media player 12. If participants were unsure what category to rate a particular video clip or if they wanted to watch it again they were allowed to watch the video clip once more if necessary.

Based on the interviews shown in the videos all subjects were required to rate each child's performance based on the CAP-II rating scale (Table 2). Each participant was required to tick only one of the categories on the rating scale that they felt most applied to each individual discussed in each of the video clips.

If participants had any further comments about any of the videos they were given the option to write these at the bottom of each of their record sheets that corresponded to each video clip (Appendix 4). At the end of the study participants were asked if they had any further comments about the videos, the CAP-II rating scale or the study as a whole. If they had additional comments the investigator noted these down on the investigators record sheet for each participant (Appendix 5).

2.3.8. Pilot study

Three subjects were required to complete the pilot study following the procedure above, to ensure that they all understood what was required of them. The results obtained from the pilot study were all similar and the previous instructions only needed 2 slight modifications made. There were only two additional things that participant 2 felt needed explained in the pilot study and these were included throughout the experiment.

When the instructions sheet was read by participants, they were told that it did not matter whether all the video clips were rated the same or differently. They were also told that they could watch the video clips twice if necessary. As no major changes were made to the experimental design, the data from the 3 participants in the pilot study were included in the final results.

2.4. Recording of Results/Statistical Analysis

Analysis was conducted for all 33 participants overall including those who did not have English as their first language. The same statistical analysis was also conducted for native English participants. This was to assess whether including non-English participants in the study would have a negative effect on the overall results.

2.4.1. Graphs

The average category that was assigned to each video clip was calculated and plotted on a bar chart. The confidence intervals were also calculated and included on the graphs as error bars. This was plotted for all participants and native English participants.

The percentage number of each category that was assigned to each video was also calculated and plotted onto separate graphs for each video. Again this was done for all participants and native-English participants.

2.4.2. Kolmogorov–Smirnov test (K-S test)

The K-S test using SPSS was used to test whether or not the data was normally distributed. This test was used as it is a non-parametric test and the data analysed in this study is non-parametric as the CAP-II rating scale is an ordinal scale and the data is continuous (Kolmogorov–Smirnov goodness-of-fit test (2003-2010). The distribution of the data was calculated first so that it can be decided what further statistical analysis can be conducted. If the value of the exact significance (2-tailed) value is > 0.05 then it is accepted that the data is normally distributed (Kinner, and Gray, 2004).

2.4.3. Spearman's Rank/Spearman's rho Correlation Coefficient

Correlation coefficient was calculated between all participants to see if the results were highly correlated. Spearman's rank using SPSS was used as the data is ordinal. This was to assess whether there was a high correlation/relationship between each participant's responses with the responses of all other participants included in the study.

The value of correlation coefficient is always between +/-1. If the value is +1 then the relationship is a perfectly positive correlation between two values but if it is -1 then the relationship is a perfectly negative correlation between two values. If the significance value is > 0.05 then it is accepted that there is not a high correlation between the two values (Barcelona Field Studies Centre, 2009).

The correlation coefficient was calculated for all 33 participants. This was also calculated for 25 native English participants. This was calculated to see if there was an effect of participant's first language. It was also to see if including non-native English participants had a negative impact on the results.

2.4.4. Cohen's Kappa Statistic

Cohen's kappa statistic was used to calculate the inter-rater reliability of the participants in the study as the main aim of the study is to assess how reliable the CAP-II rating scale is. This was calculated using an on-line kappa calculator (Randolph, 2008).

Free-marginal kappa was used as opposed to fixed-marginal kappa as Brennan and Prediger (1981) suggest using this value when the raters/participants are not forced to assign a certain number of cases to each category. The kappa value can range between -1.0 and 1.0. A kappa value of 1.0 indicates perfect inter-rater agreement above chance. A value of 0.0 indicates inter-rater reliability equal to chance. However it is usually accepted that a value of ≥ 0.7 indicates adequate inter-rater agreement (Randolph, 2008).

Cohen's kappa statistic was calculated for all 33 participants who rated all videos overall and then it was calculated for all 33 participant for each individual video clip. This was also calculated for 25 native English participants for all video clips overall and each individual video. This was calculated to see if there was an effect of participant's first language on inter-rater reliability. The results of Cohen's kappa statistic can be found in the results section, one table included all participants and the other table included native English participants.

3. Analysis of Results

3.1. Introduction: Demographics

The number of participants that were required to take part in this study initially was a total of 30 healthy participants altogether with good hearing, preferably with English as their first language. However it proved difficult to find enough participants with this requirement. Eight participants in the initial sample of 30 did not have English as their first language; however most of them were bi-lingual (7 participants) and showed a good understanding of English. As a result of this it was thought that these participants may have a negative effect on the results and therefore it would not be possible to include them in the study. However after statistical analysis was conducted with and without non-native English participants, it was found that non-native English participants did not have a negative effect on the results. So it was thought acceptable to include them in the study. In total 33 participants were willing to take part and included in the study.

All participants were healthy individuals who had no extensive experience of children who use CIs. The age of participants ranged from 19 years to 74 years of age with an average age of 29 years. The majority of participants were female, 76% and there were only 24% of participants that were male included in the study (Appendix 8).

A requirement for this study was to have all participants with English as their first language. However there were a small percentage of participants that did not have English as their first language, 24%. The majority of these participants, 21% were bi-lingual. The other 76% of participants all had English as their first language. Percentages were calculated from the information that participants included on their record sheets that were used to complete the study (Appendix 4).

All participants completed the study and none of them were excluded at any time. The results of the three participants that were included in the pilot study were also included. Changes to the instructions of the experiment were not required. However, two minor modifications were made after the pilot study was conducted.

These were as follows, participants were told that:

1. 'It does not matter whether or not the videos are all rated the same or differently.'
2. 'You are allowed to repeat the videos at least once if necessary.'

3.2. Distribution of Data

Kolmogorov–Smirnov test (K-S test) was used to analyse the distribution of the data. This statistical method was used as the data in non-parametric; the CAP-II is an ordinal, continuous rating scale. The data in this study was not normally distributed as the exact significance (2-tailed) value was $p<0.05$ for all 6 of the video clips (Appendix 9). The significance value was $P = 0.000$ for all of the video clips. Any value that is below $P = 0.05$ is accepted as not normally distributed. Consequently, non-parametric tests were appropriate.

3.3. Duration of the Study

3.3.1. All Participants

All participants were allowed to view the videos more than once, maximum two times. This was allowed in case participants were unsure what CAP-II rating to assign. It was also allowed in the instance that they may miss something that they felt may be important and help them decide what rating to give. However most participants only viewed the video clips once. As a result of this there are variations in the length of time it took participants to watch the video clips.

The length of time it took participants overall to complete the study was between 19 to 42 minutes, an average of approximately 25 minutes (24.73 minutes) altogether to watch all 6 video clips (Appendix 8). It took slightly more time for native English participants to watch them, an average of 24.8 minutes. The majority of participants overall did not watch the video clips more than once, 28 participants (Appendix 8). One participant needed to watch all the video clips twice, participant 15 (Appendix 10). Another participant, participant 2 watched 3 video clips twice, video clip 1, 3 and 5. The other 9% of participants only watched one video clip twice (Appendix 10).

Video clips 1 and 3 were repeated the most, both were repeated on average 0.09 times. Video clips 4 and 5 were repeated the same amount of times, on average 0.06 times (Table 3). Video clips 2 and 6 were repeated the least, on average 0.03 times (Table 3).

Table 3: The average number of repetitions that all participants required to watch each video clip.

Video Number	Average Number of Repeats for All Participants	Average Number of Repeats Native English Participants
1	0.09	0.04
2	0.03	0.04
3	0.09	0.08
4	0.06	0.08
5	0.06	0.04
6	0.03	0.04

3.3.2. Native English Participants

The length of time it took participants excluding those that do not have English as their first language to watch all 6 video clips was between 20 minutes and 42 minutes, an average of 25 minutes also (24.8 minutes) (Appendix 11). This is only slightly longer than the results obtained including non-English participants. 88% of participants did not need the video clips repeated when non-English participants were excluded. This is only 3% more than if non-English participants are included in the study so it seems unlikely to affect the results overall.

The participant that needed to watch all video clips again was a participant who had English as their first language as they were included in the results of Native English participants (Appendix 10). The other 8% of participants needed the video clips repeated at least once (Appendix 11).

The average number of repeats of native English participants was slightly less for video clip 1, 3, and 5. Video clip 1 was repeated on average 0.04 times as opposed to 0.09 when non-English participants were included (Table 3). Video clip 3 was repeated on average 0.08 times as opposed to 0.09 times and Video clip 5 was repeated 0.04 times as opposed to 0.06 times (Table 3).

However video clips, 2, 4 and 6 were repeated more when non-English participants were excluded. Video 2 was watched on average 0.03 times whereas when non-English participants were excluded it was watched on average 0.04 times (Table 3).

Video clip 4 was watched on average 0.08 times as opposed to 0.06 times when non-English participants were excluded (Table 3). Video 6 was also watched more when non-English participants were excluded, on average 0.04 times as opposed to 0.03 times (Table 3). These results including non-English participants in the study did not have a significant difference on the number of repetitions of the 6 videos clips.

3.4. CAP-II Scores Assigned to Each Video clip

3.4.1. Ratings Assigned Overall and by Native English Participants

3.4.1.1. Video Clip 1

The minimum value given to video clip 1 was category 5 and the maximum given was 7 (Appendix 12). Video clip 1 was given a mean CAP-II score of 6.70, (category 7) overall (Figure 4) (Appendix 12). However, native English participants had a mean CAP-II rating of 6.76, (category 7) (Figure 5) (Appendix 13). These results show that these ratings are very similar although the mean is slightly higher for native English participants.

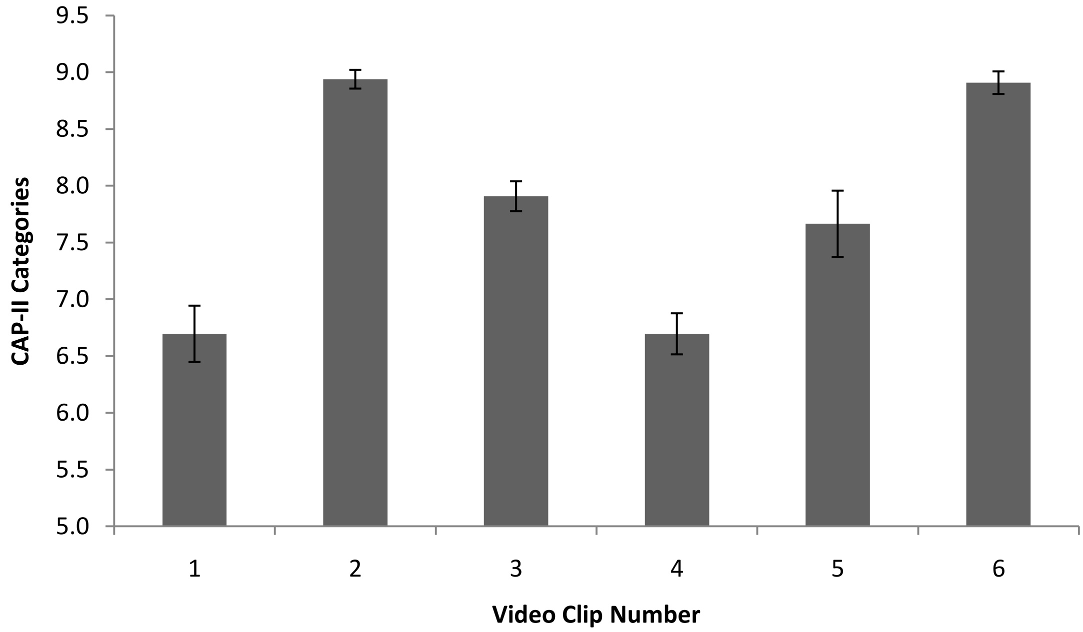

Figure 4: The mean CAP-II category assigned to each video clip by all 33 participants included in the study. The error bars represent the upper and lower 95% confidence intervals (p = 0.05).

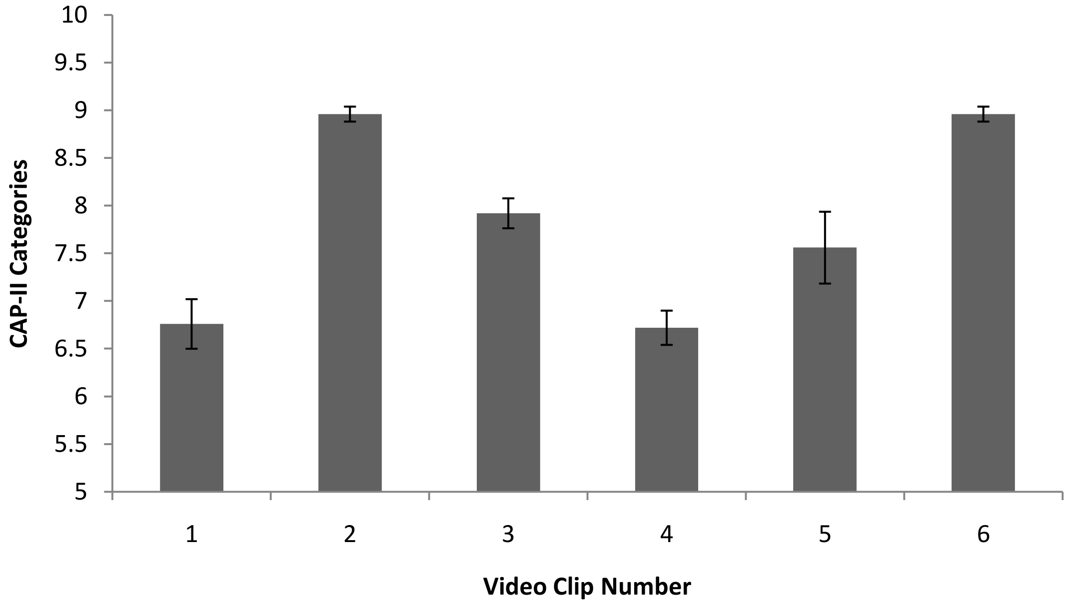

Figure 5: The mean CAP-II assigned to each video clip by native English participants. The error bars represent the upper and lower 95% confidence intervals (p = 0.05).

The standard deviation of the mean was 0.728 overall and 0.663 for native English participants (Appendix 12 and 13). The confidence intervals overall was p = 0.248 (Appendix 14). For native English participants it was p = 0.26 (Appendix 15).

This is greater than p = 0.05 which indicates that the ratings assigned to this video clip are not statistically different. Any value that is > 0.05 is not statistically significant (Rowans University, 2010). The 95% confidence intervals for video clip 1 were 6.70 and 6.44 overall (Figure 4). For native English participants they were 7.03 and 6.49 (Figure 5).

A total percentage of 81.82% overall rated this video clip a CAP-II score of 7, 15.15% rated it a CAP-II score of 5 and the other 3.03% rated this video a CAP-II score of 6 (Figure 6). There was a slight difference in the percentage number of native English participants that assigned each category to video clip 1. 12% rated this video a CAP-II score of 5, 84% of participants give a score of 7 and the other 4% 6 (Figure 6). 2.18% more native English participants rated this video clip category 7 however this is not a big difference.

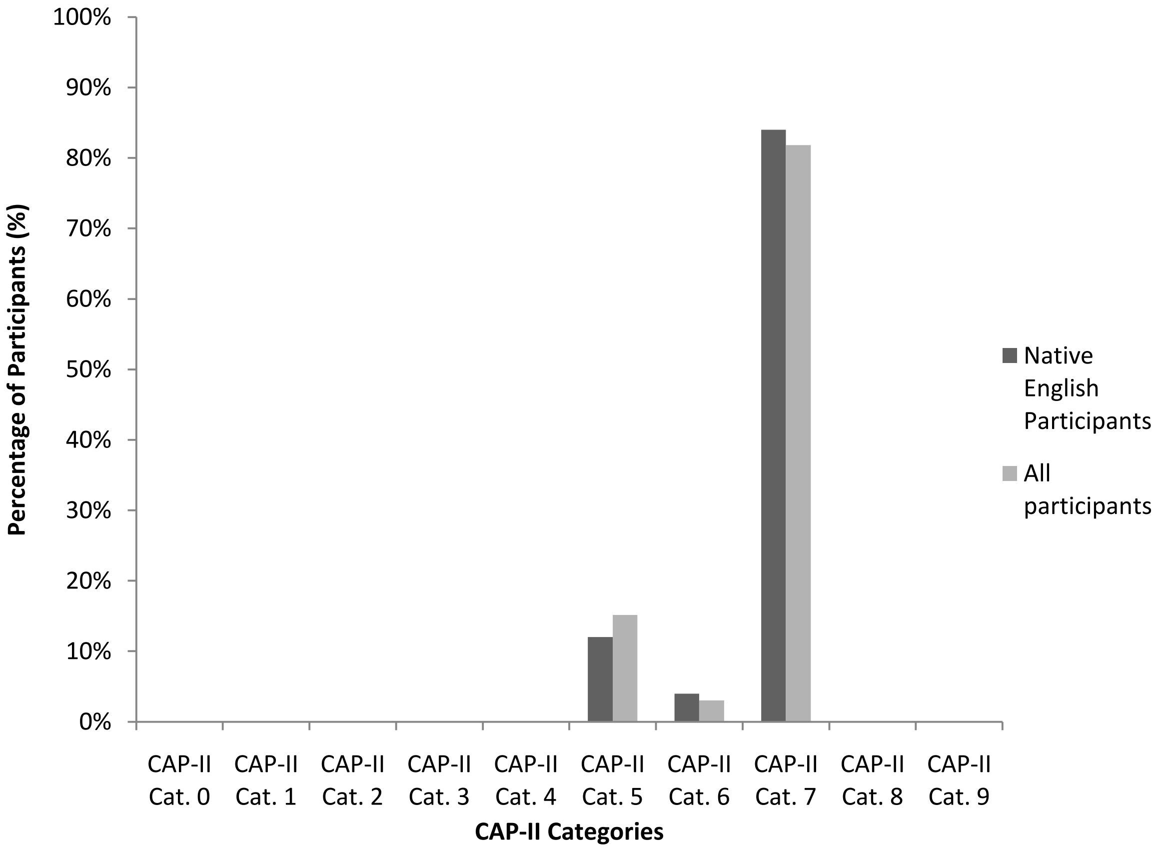

Figure 6: The percentage number participants that assigned each CAP-II category to video clip number 1. (Dark grey represents native English participants, light grey represents all participants).

3.4.1.2. Video Clip 2

The minimum category assigned to video clip 2 was 8 and the maximum was category 9 (Appendix 12). This is the maximum category possible with the CAP-II rating scale (Table 2). Video clip 2 was given a mean score of 8.94, (category 9) overall (Figure 4) and a mean of 8.96 (category 9) by native English participants (Figure 5) (Appendix 13).

The standard deviation was 0.242 for native English participants it was 0.20. The confidence intervals were p = 0.083 overall (Appendix 14) and for native English participants they were p = 0.08 (Appendix 15). This is > 0.05 which indicates that the ratings assigned to this video clip are not statistically Significant. The 95% confidence intervals overall were 9.03 and 8.85 for video clip 2 (Figure 4). For native English participants they were 9.04 and 8.88 (Figure 5).

Only a small percentage of all participants assigned category 8, 6.06% of participants (Figure 7). An even smaller percentage, 4% of native English participants give category 8 (Figure 7). A large percentage overall, 93.94% assigned the maximum CAP-II score, category 9 (Figure 7). The majority of native English participants, 96% give a CAP-II rating of 9 (Figure 7). This is only 2.06% more than if non-native English participants are included so there is not a large difference in the ratings assigned (Figure 7).

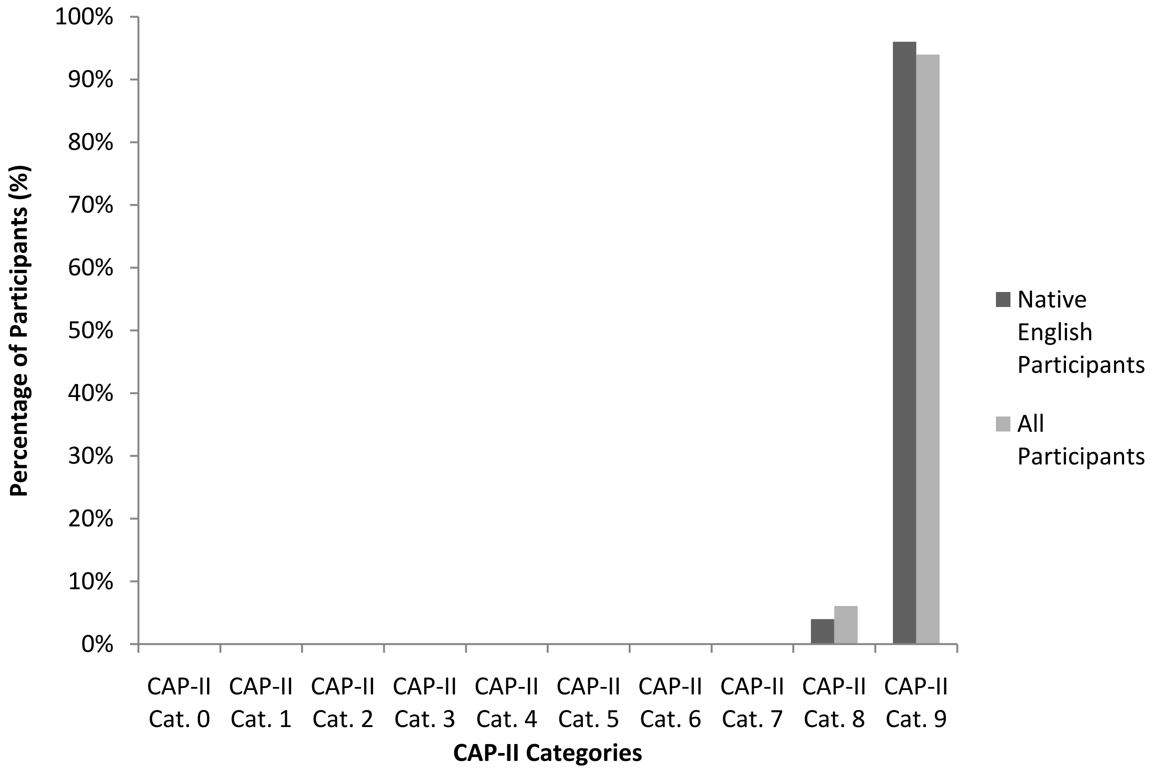

Figure 7: The percentage number of participants that assigned each CAP-II category to video clip 2. (Dark grey represents native English participants, light grey represents all participants).

3.4.1.3. Video Clip 3

The minimum CAP-II score assigned to video clip 3 was category 6 and the maximum was 8 (Appendix 12). Video clip 3 was assigned a mean category of 7.91, (Category 8) overall (Figure 4) (Appendix 12). Non-native English participants had and a mean category of 7.92, (category 8) (Figure 5) (Appendix 13). The standard deviation was 0.384 overall and was 0.40 for native English participants (Appendix 12 and 13).

The 95% confidence intervals overall were 8.05 and 7.77 this is the equivalent of p = 0.131 of the mean (Figure 4). For non-native English participants they were 8.09 and 7.75, p = 0.16 (Figure 5). The p value for this video clip was p>0.05 which indicates that the ratings assigned to this video clip are not statistically significant.

Only a small percentage of all participants allocated a category of 6 and 7, 3.03% for each of these categories (Figure 8). However, none of the native English participants rated category 7 and 4% of them give a CAP-II score of 6 (Figure 8). The majority participants overall, 93.94% rated this video clip category 8 (Figure 8).

A large percentage, 96% of native English participants give a CAP-II score of 8 (Figure 8). This is a difference of 2.06% (Figure 8).

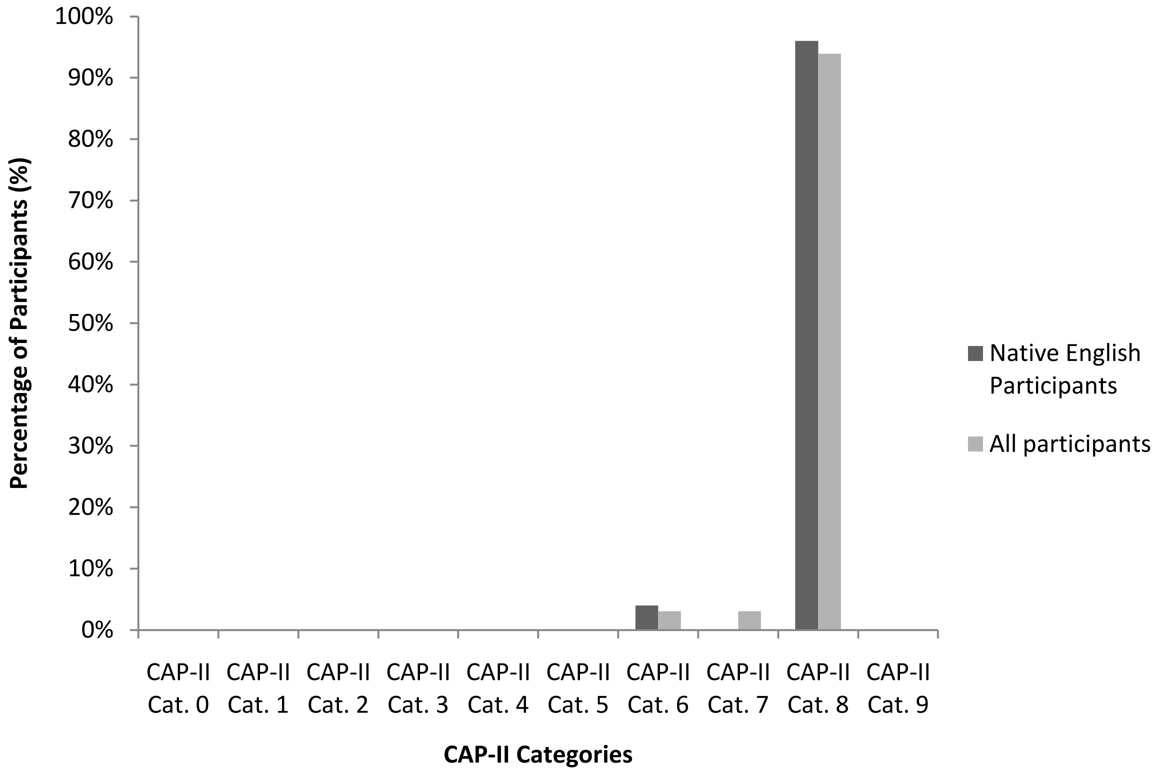

Figure 8: The percentage number of participants that assigned each CAP-II category to video clip 3. (Dark grey represents native English participants, light grey represents all participants).

3.4.1.4. Video Clip 4

All participants assigned a minimum CAP-II rating of 5 and a maximum of 7 to video clip 4. However the minimum CAP-II assigned by native English participants was CAP-II 6. Video clip 4 was given a mean rating of 6.69, (Category 7) overall Figure 4) (Appendix 12). Native English participants had a mean rating of 6.72 (category 7) (Figure 5) (Appendix 13).

The standard deviation was 0.529 and the 95% confidence intervals were 6.88 and 6.51, $p = 0.181$ overall (Figure 4) (Appendix 14). The standard deviation for native English participants was 0.458 and the confidence intervals were 6.91 and 6.53, $p = 0.18$ (Figure 5) (Appendix 15). The ratings assigned to this video clip are not statistically different as $p>0.05$.

3.03% rated category 5, 72.73% category 7 and the other 24.24% overall give category 6 (Figure 9). 28% of Native English participants assigned category 6 and 72% assigned category 7 (Figure 9). Approximately the same percentage of participants assigned category 7 both overall and native English participants (Figure 9). None of the native English participants give a rating of category 5 (Figure 9). However when non-native English participants are included there is a small percentage that give a category of 5, 3.03% (Figure 9).

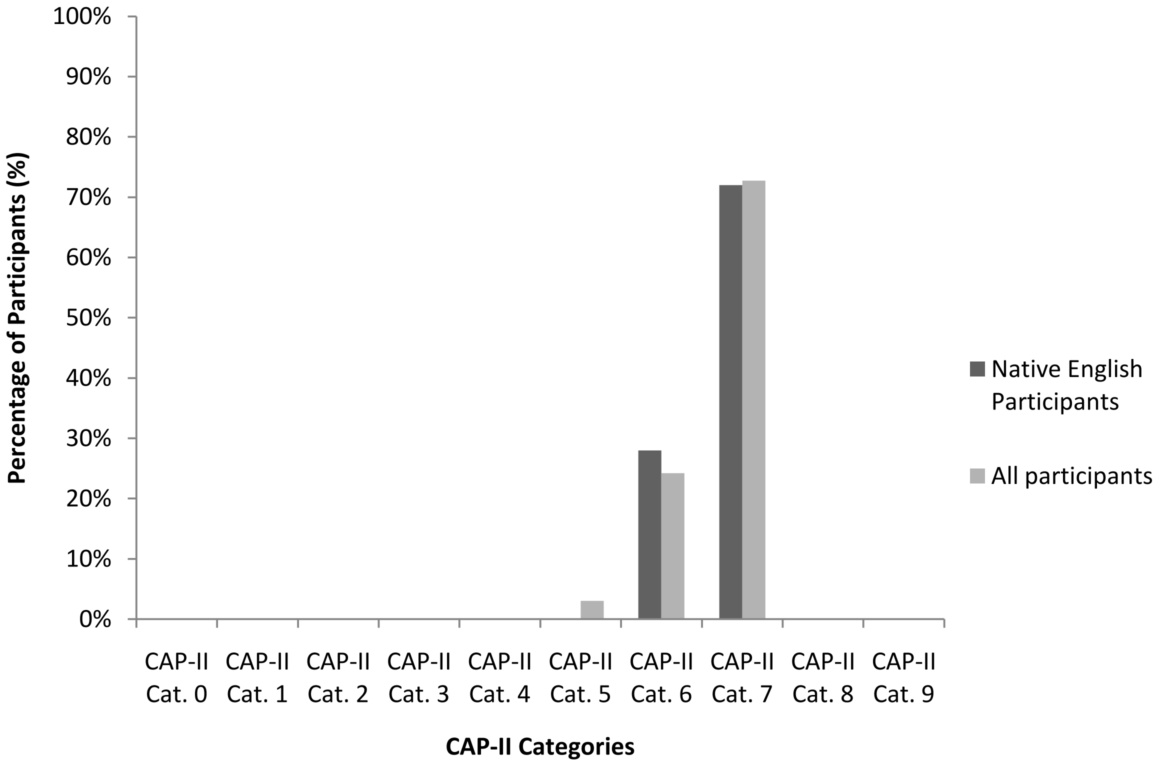

Figure 9: The percentage number of participants that assigned each CAP-II category to video clip 4. (Dark grey represents native English participants, light grey represents all participants).

3.4.1.5. Video Clip 5

The minimum CAP-II rating given for video clip 5 was 4 and the maximum was 8 (Appendix 12). This video clip had the largest variation in categories assigned, 4 different categories. Video clip 5 was given a mean rating of 7.67, (Category 8) overall (Figure 4) (Appendix 12) and a mean of 7.56, (category 8) by native English participants (Figure 5) (Appendix 13). The standard deviation was 0.854 and the confidence intervals were p = 0.291 overall (Appendix 14).

For native English participants the standard deviation was 0.961 and the confidence intervals were p = 0.38 (Appendix 15). There is no significant difference in the ratings assigned p>0.05.

The 95% confidence intervals were 7.97 and 7.36 overall (Figure 4). For native English participants they were 7.96 and 7.16 (Figure 5). Only a very small percentage of participants overall, 3.03% assigned a CAP-II rating of 4 (Figure 10). No participants assigned a CAP-II of 5, 6.06% assigned a CAP-II of 6, 9.09% assigned category 7 and the majority of all participants, 81.82% assigned a CAP-II of 8 (Figure 10).

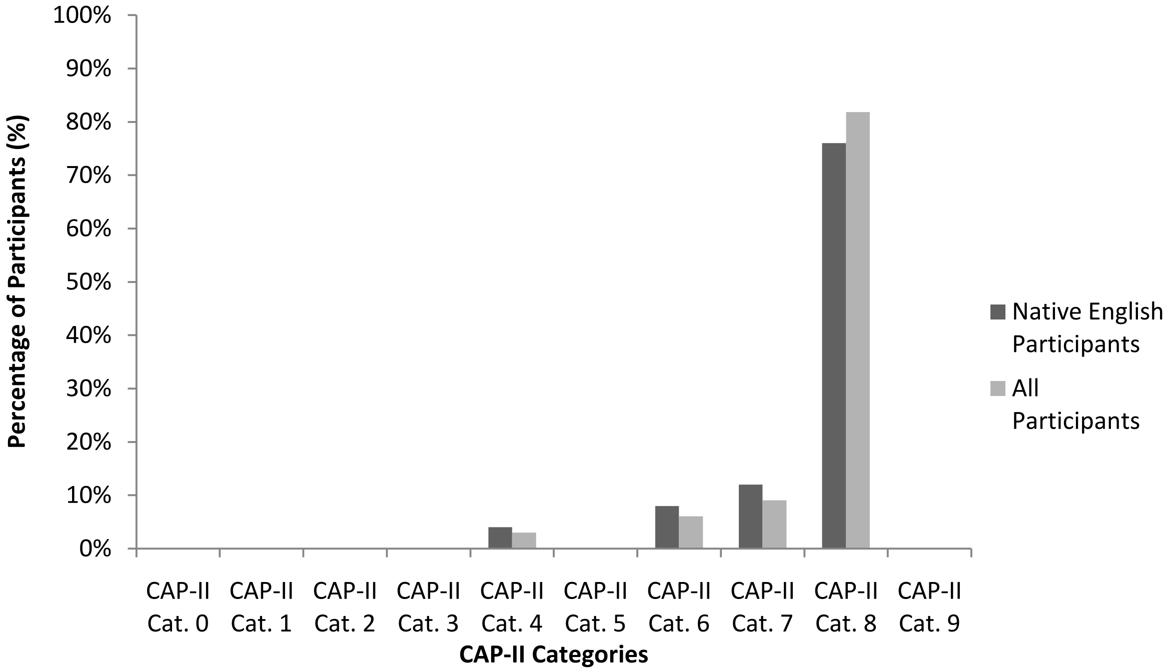

Figure 10: The percentage number of participants that assigned each CAP-II category to video clip 5. (Dark grey represents native English participants, light grey represents all participants).

4% of native English participants give a CAP-II score of 4, 0% give 5, 8% give a CAP-II score of 6, 12% give a 7 and the largest percentage, 76% give a rating of 8 (Figure 10). When non-native English participants were included a higher percentage, 5.82% of participants give a rating of 8 to video clip 5 (Figure 10).

3.4.1.6. Video Clip 6

The maximum CAP-II rating given to video clip 6 was the maximum possible category, category 9 and the minimum was 8 (Appendix 12). Video Clip 6 was assigned a mean CAP-II score of 8.91, (Category 9) overall (Figure 4) (Appendix 12). For native English participants a mean CAP-II rating of 8.96 (category 9) was given (Figure 5) (Appendix 13). The standard deviation was 0.292 and the confidence interval was p = 0.100 (Appendix 14) overall. For native English participants the standard deviation was 0.20 and the confidence interval was p = 0.08 (Appendix 15). There is no significant difference in the ratings assigned as $p>0.05$.

The 95% confidence intervals were 9.01 and 8.81 overall (Figure 4). For native English participants the 95% confidence intervals were 9.043 and 8.88 (Figure 5). 12.12% of all participants assigned a CAP-II category 8 and the other 87.88% assigned the maximum possible category, category 9 (Figure 11). 92% of native English participants assigned category 9 (Figure 11). This is 4.12% more than the results obtained when non-native English participants are included (Figure 11). 8% of native English participants assigned category 8 (Figure 11).

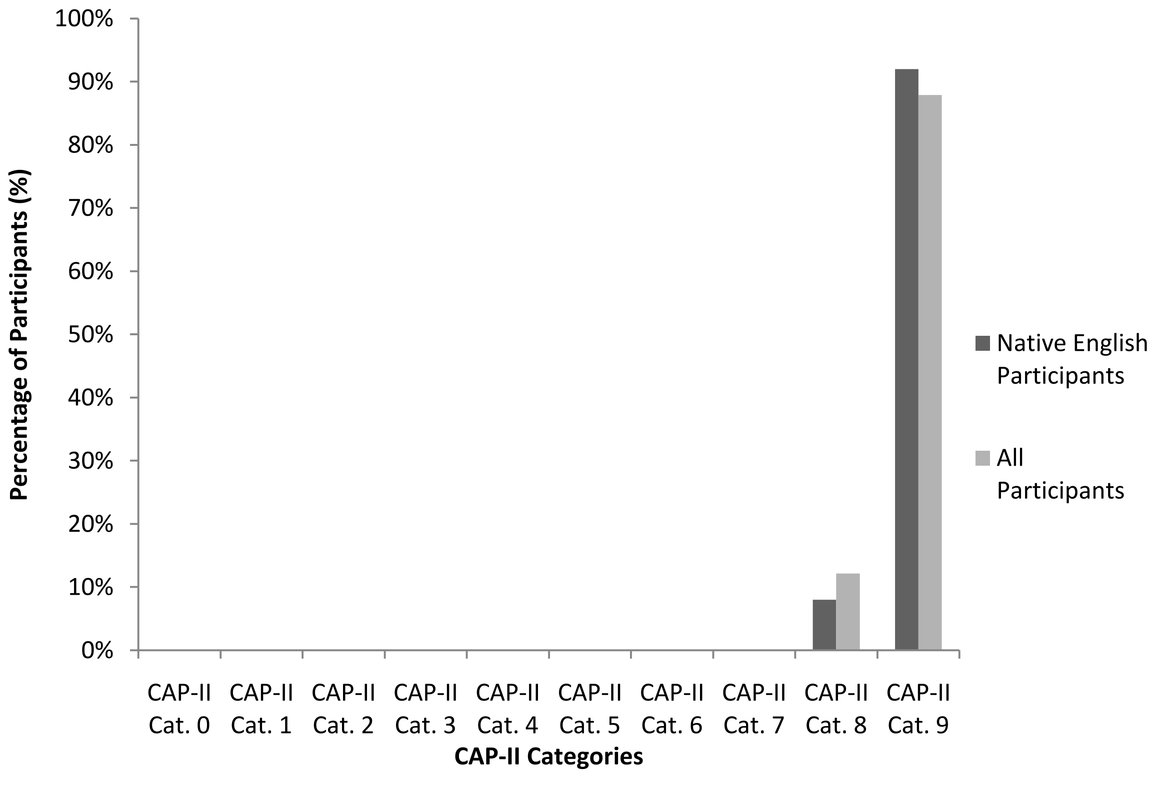

Figure 11: The percentage number of participants that assigned each CAP-II category to video clip 6. (Dark grey represents native English participants, light grey represents all participants).

Overall video clip 2 and 6 were assigned the maximum possible category, category 9 more than the other video clips (Figure 7 and 11). Video clip 2 achieved the highest percentage of participants rating a CAP-II score of 9 with 93.94% of all participants assigning this category (Figure 7).

In general there is not much variation in the ratings assigned by participants overall and when non-English participants are included in the study. As a result non-English participants' responses are included in the final analysis of the data.

3.5. Analysing Inter-Rater Reliability

3.5.1. Cohen's Kappa Statistic

Cohen's Kappa statistic was calculated for each individual video which included the CAP-II ratings assigned by all 33 participants and native English participants. This was calculated to assess the inter-rater reliability between all of the participants for each video. It was also calculated with only native English participants to see if including non-native English participants had a negative effect on the results. Cohen's Kappa statistic was also calculated for all of the videos overall to assess the inter-rater reliability of participants on the whole.

Participants were not forced to assign a certain number of cases to each category in the CAP-II rating scale so the result of the free-marginal Kappa was used as suggested by Brennan and Prediger (1981) and the overall percentage of agreement PO value was used (Table 4). Brennan and Prediger (1981) state 0.7 agreement is considered good. Any value that is ≥ 0.7 is accepted as adequate and the inter-rater reliability if probably not due to chance.

Table 4: The Cohen's kappa statistic calculated for each individual video and all 6 video clips overall for all 33 participants included in the study.

	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	All Videos
Percentage of Overall Agreement PO (%)	0.684	0.883	0.881	0.591	0.672	0.780	0.748
Fixed-Marginal Kappa	-0.031	-0.031	-0.031	-0.031	-0.031	-0.031	0.651
Free-Marginal Kappa	0.649	0.870	0.867	0.545	0.636	0.756	0.720

There was high inter-rater agreement for videos 2, 3 and 6 overall as all the values are > 0.7 (Table 4). There was also high inter-rater agreement between native English participants for these 3 video clips and video 7. Any value that is > 0.7 is accepted as adequate and the inter-rater reliability is probably not due to chance. Videos 2 and 3 had the highest inter-rater reliability values and the values were similar overall. The percentage of overall agreement was PO = 0.883% for video 2 and it was PO = 0.881% for video 3 (Table 4).

For native English participants a similar finding was found although video clips 2 and 3 had the same percentage of overall agreement, PO = 0.92% (Table 5). The free-marginal kappa value was also very high overall, for video 2 was 0.869 and for video 3 it was 0.867 (Table 4). For native English participants it was 0.911 for both of these videos (Table 5). When non-English participants were included this value was slightly lower for both video clips 2 and 3 (Table 5). The overall percentage of agreement was a difference of PO = 0.04% (Table 4 and Table 5). Including non-English participants does not have a significant impact on the inter-rater reliability values.

Table 5: The Cohen's kappa statistic calculated for each individual video and all 6 video clips native English participants.

	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	All Videos
Percentage of overall agreement PO (%)	0.710	0.920	0.920	0.580	0.583	0.847	0.760
Fixed-Marginal Kappa	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	0.668
Free-Marginal Kappa	0.678	0.911	0.911	0.533	0.537	0.830	0.733

Video 1 also had high inter-rater reliability overall although the percentage of overall agreement was slightly < 0.7 but it was only 0.02% below this accepted value. However for native English participants it was > 0.7 the value was PO = 0.71. This is this is 0.03% higher than the results found overall for video 1. The free-marginal kappa overall was also < 0.7, but only by a value 0.05 (Table 4). For native English participants it was slightly higher, PO = 0.678. This is also a difference of 0.03 (Table 5).

Video 5 had a similar inter-rater reliability value overall as video 1 only it was 0.01 below the overall percentage and the free-marginal kappa value of video 1 (Table 4). For native English participants the overall percentage agreement for video clip 5 was PO = 0.58% (Table 5). However the results obtained for participants overall was slightly higher for this video, PO = 0.672% (Table 4). This is a difference of 0.09% this is highest difference between the inter-rater reliability values for all of the video clips. It also shows including non-English participants does not have a negative effect on inter-rater reliability. The free-marginal kappa value was also higher overall, 0.636 as opposed to 0.537 when non-native English participants are excluded (Table 4 and 5).

The video that had the lowest inter-rater reliability was video 4 overall. The percentage of overall agreement PO = 0.591 and the free-marginal Kappa overall = 0.545 (Table 4). Even though this value is < 0.7 it is still considerably high and is very close to 0.7. This value is only 0.1 below the recommended accepted value for accepting high inter-rater reliability above chance.

However the overall percentage for video 4 is slightly lower, 0.01% for native English participants as the PO = 0.58 (Table 5). This shows that for this particular video clip that inter-rater reliability is higher when non-English participants are included. The free-marginal kappa value is also slightly less overall; the value is 0.55 as opposed to 0.53 for native English participants (Table 4 and 5).

Video clip 6 had an overall percentage agreement of PO = 0.847% for native English speakers (Table 5). This value is 0.07% higher than the overall value as the PO = 0.78% for all participants (Table 4). The free-marginal kappa value was also slightly higher for native English participants. For native English participants the value is 0.83 whereas the value is 0.756 overall (Table 4 and Table 5).

However the inter-rater reliability of all the videos overall was relatively high. The percentage of overall agreement PO = 0.748% and the free-marginal kappa = 0.720 overall (Table 4). Native English participants had slightly higher inter-rater reliability values overall. The percentage of overall agreement PO = 0.76% and the free marginal kappa value was 0.733. This is only 0.01% higher which is not considerably high.

The high inter-rater reliability indicates that participants did not rate all the videos in the study similarly just by chance alone. This value indicates that there is inter-rater reliability perfectly above chance. A difference of 0.01% between the results overall and native English participants indicates that it is unlikely that native use of English is an important factor in determining inter-rater reliability with the CAP-II rating scale.

Overall including non-native English participants in the study did not have a significant negative effect on the inter-rater reliability values overall. All the results indicate that there is high inter-rater reliability in this study both with and without non-native English participants. In some instances however, especially with video clips 5 and 6 including non-English participants increased the inter-rater reliability scores for these two videos.

3.5.2. Spearman's Rank/Spearman's rho Correlation Coefficient

The correlation coefficient was calculated overall to assess the relationship between participants and their responses. The relationship between each individual participant with all of the other participants in the study is included in Appendix 16.

The correlation coefficient calculation revealed that there is a very strong positive relationship between all participants. If the correlation coefficient value is +1 then this indicates a perfectly positive correlation between two values.

Table 6 shows the mean relationship between each participant with all the other 32 participants. Overall the majority of participants had a high correlation coefficient; the mean value was > 0.75 (Table 6). For native English participants it was also high, a mean value of > 0.739 (Table 7). However there were a few participants overall, 6 participants whose mean was slightly less than this value although it is not considerably less as they all have a value of > 0.62 (Table 6). There were only 2 participants in the native English group that had a value < 0.739 although it was still high, > 0.631 (Table 7). These results indicate a high positive relationship on average between all participants and native English participants.

When each individual participant was compared with all the other participants the majority of them had a high correlation overall ($p<0.05$) (Appendix 16). However there were a few individual participants that had a low correlation with some of the participants ($p>0.05$) (Appendix 16).

Overall the result revealed that there is a high correlation between all participants both when non-English participants are included and also when they are not. This indicates that including non-English participants does not have a negative effect on the results obtained overall.

Table 6: The mean correlation coefficient and the mean significance value of all participants.

Participants	**Mean Correlation Coefficient**	**Mean Significance (2-tailed)**	**Number**
1	0.765	0.022	6
2	0.666	0.079	6
3	0.763	0.034	6
4	0.763	0.034	6
5	0.763	0.034	6
6	0.668	0.059	6
7	0.765	0.022	6
8	0.763	0.034	6
9	0.763	0.034	6
10	0.718	0.108	6
11	0.765	0.022	6
12	0.763	0.034	6
13	0.754	0.056	6
14	0.763	0.034	6
15	0.763	0.034	6
16	0.765	0.022	6
17	0.763	0.034	6
18	0.763	0.034	6
19	0.690	0.060	6
20	0.759	0.021	6
21	0.759	0.021	6
22	0.785	0.024	6
23	0.765	0.022	6
24	0.763	0.034	6
25	0.763	0.034	6
26	0.685	0.051	6
27	0.763	0.034	6
28	0.763	0.034	6
29	0.763	0.034	6

30	0.623	0.202	6
31	0.763	0.034	6
32	0.664	0.157	6
33	0.763	0.034	6

Table 7: The mean correlation coefficient and mean significance value of native English participants.

Participants	Mean Correlation Coefficient	Mean Significance (2-tailed)	Number
1	0.765	0.027	6
2	0.763	0.038	6
3	0.763	0.038	6
4	0.763	0.038	6
5	0.765	0.027	6
6	0.763	0.038	6
7	0.739	0.089	6
8	0.765	0.027	6
9	0.781	0.049	6
10	0.763	0.038	6
11	0.763	0.038	6
12	0.765	0.027	6
13	0.763	0.038	6
14	0.749	0.050	6
15	0.747	0.025	6
16	0.747	0.025	6
17	0.793	0.022	6
18	0.765	0.027	6
19	0.763	0.038	6
20	0.763	0.038	6
21	0.763	0.038	6
22	0.631	0.196	6
23	0.763	0.038	6
24	0.682	0.140	6
25	0.763	0.038	6

4. Discussion

4.1. Introduction

The aims of the present study outlined in chapter 2 were as follows 1) to evaluate the inter-rater reliability of the CAP-II rating scale for future assessments of children with CIs and 2) to determine whether the two additional categories are sufficient for assessing young children with CIs and those that are bilaterally fitted whilst avoiding ceiling effects.

The literature review strongly indicated that the original CAP rating scale has very high inter-rater reliability. Archbold *et al*., (1998) conducted a study investigating the inter-rater reliability of the original CAP scale. The analysis of the scores assigned by local teachers of the deaf who see the children frequently and by those who did not see the children very often revealed very high inter-rater reliability with a correlation coefficient value of 0.97 (Archbold *et al*., 1998).

The original CAP rating scale has been used widely by professionals and parents worldwide as it is a very reliable and valid assessment method (Nikolopoulos *et al*., 1999). However there has been evidence that the current CAP rating scale needs to be modified and further developed as ceilings effects have become apparent (Beadle *et al*., 2005; Wu *et al.,* 2007). It is limited in that it cannot measure the recent benefits that are now achievable with CIs and bilateral fittings i.e. 'use of the telephone with an unknown speaker in unpredictable context.' So it has been proposed that two new categories should be added to the original CAP scale to increase the number of benefits that can be assessed (CAP-II). However inter-rater reliability needs to be assessed to determine whether the two new proposed categories are suitable for assessing these benefits. As a result the main purpose of this study was to investigate the inter-rater reliability of the CAP-II rating scale. The CAP-II rating scale needs to be valid and reliable in order to be an effective assessment method and to determine internal consistency.

This study revealed high inter-rater reliability with the CAP-II rating scale which is in keeping with Archbold's *et al*., (1998) findings that inter-rater reliability is high for the original CAP. There was a high correlation coefficient between all the raters for the CAP-II, although the values are slightly lower than Archbold *et al.,* (1998) found with the original CAP rating scale. On average the majority of raters had a mean correlation coefficient of > 0.75 (Table 6 and 7). Cohen's kappa values were also high for all of the video clips indicating high inter-rater reliability.

This study also indicated that the two additional categories are sufficient for assessing young children with CIs and those that are bilaterally fitted, as it covers the additional benefits that children are now able to receive with their CIs. The new CAP-II rating scale did not result in ceiling effects as all the raters did not assign the maximum category, category 9 (Table 2) to each child discussed in the video clips. Only two of the video clips out of all 6 were rated a CAP-II score of 9, video clips 2 and 6. However not all raters assigned category 9 to these two video clips some participants rated them a score of 8.

4.2. Duration of the Study

The length of time it took participants to complete the study was investigated in order to account for any repetitions of the video clips if they were necessary. There was some variation in the length of time it took participants to watch the video clips because some participants needed to repeat some of the video clips whereas other participants did not. Also some participants wrote a number of comments on their record sheet throughout the study which took time whereas others did not.

One participant, participant 15, had to watch all video clips twice as they watched them in the wrong order and labelled the sheets wrong, resulting in the wrong ratings assigned to each video clip (Appendix 10 and 18). As a result of this they had to repeat the study, although this only happened the once. Analysis of the duration of the experiment was conducted both with and without participants who had English as their first language. This was to assess whether or not there would be an effect of first language on the results.

However it was found that participant 15 who needed all the video clips repeated had English as their first language. There was very little difference in the duration of the study both with and without non-native English participants so language was proven not to be an effect on the duration of the study. This indicates that the ratings could be undertaken by fluent, non-native English speakers.

On average video clips 1 and 3 were repeated the most (Table 3). The reason for this was possibly because participants found it more difficult to rate the child based on what was discussed in the video clips, the responses of the interviewees and the way in which the interviewer asked the questions. However some participants found that video clip 3 was one of the most clear video clips. Some participants struggled to rate video clip 1 as the child could do level 7 but not level 6 (Appendix 18). Category 6 is easier than category 7 so they were unsure whether to rate 5 or 7. Therefore it took longer for some participants to decide what category to assign to video clip 1 (Appendix 18).

The two video clips that had the least repetitions were video clips 2 and 6 (Table 3). These two video clips were assigned the highest category, category 9 the most (Figure 7 and 11). These two video clips also had the least additional comments given which would imply that the majority of participants found that these two video clips were very clear (Appendix 18). It was also very clear that the child in each of these video clips could do category 9. As a result of this participants were able to quickly decide which rating to give therefore most people did not required these two video clips to be repeated.

4.3. CAP-II Scores Assigned to Each Video clip

The majority of the video clips were given a difference of 3 CAP-II categories with a high percentage of participants giving the same rating and few participants rating one or two categories above or below the most common rating given (Figure 6 to 11). For example video clip 2 and 6 were given the highest ratings out of all of the video clips. The majority of raters assigned category 9 to these video clips although there were a few participants who assigned category 8 (Figure 7 and 11).

Similar findings were found for the other video clips although the CAP-II scores assigned varied and the majority of the other video clips were assigned lower categories. This illustrates that ceiling effects with the CAP-II rating scale were not apparent as none of the video clips were rated the highest category, category 9 by all participants included in the study. The two video clips that were rated category 9 were also not rated this category by all of the participants in the study.

Some video clips were more difficult to judge than others which resulted in a lot of variation in the ratings assigned. For example in video clip 1 the majority of participants rated this video clip a score of 7 (Figure 6). Hardly any participants assigned a score of 6 and quite a large percentage assigned a score of 5 (Figure 5 and 6). The reason why participants found this video so difficult to judge was because the child could do category 5 and 7 but not category 6. Participants were unsure whether to rate the lowest level that the child could do as a result of this or the highest level. This meant that there were a considerable percentage of participants giving scores of 5 and 7.

This indicates that although the CAP-II is a good reliable assessment method other assessment methods need to be used in conjunction with it as recommended by Beadle *et al.,* (2005). The reason for this is because children may be rated higher in their abilities than what they are actually capable of. The CAP-II is a hierarchy of skills meaning that children should be able to do all levels of the scale up until the rating they are assigned. However this may not be the case as indicated by video clip 1 as the child is given a CAP score of 7 by the majority of participants yet they cannot do category 6.

It is unusual that the child discussed in video clip 1 could do category 7 but not do category 6. The reason this is odd is because the child cannot understand conversation without lip-reading yet they can talk on the telephone. This is an extraordinary finding as a child cannot lip-read whilst using the telephone and usually most people with a HI struggle the most on the telephone as the telephone can distort some of the speech frequencies.

Video clip 5 had the most variation in the ratings assigned there was a difference of 4 CAP-II ratings given (Figure 10). The CAP-II score ranged from category 4 to category 8, with no participants assigning category 5 (Figure 10). None of the participants assigned category 9 to this video clip either. The interviewee did not know whether or not the child could use the telephone with an unknown speaker due to their age so this had not been investigated. There was only one participant, participant 32 that assigned category 4 to video clip 5 (Appendix 8). This could be due to the participant assigning the wrong category due to missing something essential discussed in the video clip as she stopped it just before it had finished.

4.4. Limitations based on Interview Technique

The majority of participants found that the video clips were clear and well discussed (Appendix 18). They found that the interviewer asked the questions well and that the interviewees' responses were clear. However there were two participants, who found that the interviewer could be quite leading at times when asking the questions therefore influencing the responses she was obtaining (Appendix 18).

A few participants found that the interviewer asked the interviewee different questions when referring to category 9. For example the interviewer asked the interviewee discussing the child in video clip 3 about the child using the telephone in noise when referring to category 9 but this is not specified on the CAP-II rating scale (Appendix 1 and 18). However when the interviewer asked the interviewee about the child in video clip 6 she asked about the child using the telephone in quiet when referring to category 9 (Appendix 1 and 18). This is a different situation to what was asked in video clip 3 (Appendix 18). This may have resulted in none of the participants assigning a CAP-II score of 9 to video clip 3. As the question was asked incorrectly and a more difficult task was referred to in video clip 3 as opposed to the child not being able to do category 9.

A few participants found that the interviewer missed out some of the categories in the rating scale and then went back to the categories missed out, in some of the video clips (Appendix 18). I.e. the interviewer asked about category 6, then category 8 and then went back to category 7. Questioning was found to be unpredictable at times by participants and the questions were not asked in a controlled way.

This was found especially in video clips 1 and 5, the questions were not asked in order so participants found these video clips more difficult to rate (Appendix 18). Therefore this resulted in variations in the ratings assigned. This could also account for why lower ratings were given to some of the video clips as people may assign the last category that was discussed in the video clips that the child could do as opposed to the hardest level discussed throughout the video clips.

Some participants also found that one of the interviewees in video clip 3 based a lot of her responses of what the child says to her as opposed to her own observations (Appendix 18). There could be bias in the responses given and therefore the ratings assigned may not be a true reflection of how the child is actually doing. What the child has told the interviewee may be incorrect. The child may not be able to do as well as they say they are doing.

Two participants queried the terminology in video clip 4 as they were unsure what the interviewee meant when she was talking about 'decrements' (Appendix 18). This could have had a negative effect on the ratings assigned to video clip 4. Some participants may not have given correct ratings as they were unsure what the interviewee meant by decrements. They may have also been too embarrassed to ask what it meant, meaning that more people in the study could have been unsure about the terminology than was revealed in the study.

4.5. Inter-Rater Reliability

Cohen's kappa statistic and correlation coefficients were used to assess the inter-rater reliability of the CAP-II rating scale. Statistical analysis was calculated for all participants, which includes native English participants and non-Native English participants. It was also calculated for only native English participants to see if native language had an effect on inter-rater reliability. The results of Cohen's kappa statistic for all the video clips overall indicated that the results from native English speakers and those including non-native English participants were very similar and there was not a significant difference in the ratings assigned. There was only a 0.1% difference in the percentage of overall agreement between native English participants and all participants for all 6 of the video clips when Cohen's kappa was calculated (Table 4 and 5).

This indicates that fluent non-native English speakers can use the CAP-II rating scale so the results of these participants are included in the study. This is a positive finding as the original CAP rating scale is used worldwide to assess young children (Wu *et al*., 2005; Nikolopolous *et al.,* 1999a). If fluent non-native English participants can use the CAP-II rating scale perhaps in future it can be translated into other languages and also used worldwide like the original CAP scale (Nikolopolous *et al.,* 1999a).

Although statistical analysis was conducted for both of these groups, all participants and native English participants the results in this section will be quoted from all of the participants combined. Cohen's Kappa statistic was used as the most reliable result between native English raters and non-native English raters as it corrects chance of agreement (Nicholas *et al.,* 1999). Cordes (1994) states that percent agreement alone does not account for reliability as it is very difficult to determine on what basis percent agreement estimates are calculated so more researchers are using Cohen's kappa statistic (Nicholas *et al*., 1999).

High inter-rater reliability was found between all participants as the Cohen's kappa value was high for all 6 video clips (Table 4 and 5) and the correlation coefficient was also high between all participants (Appendix 16 and 17). Video clip 4 had the lowest value when Cohen's kappa statistic was calculated this was probably because participants found that this was the most difficult video clip to judge (Table 4 and 5).

Video clips 2, 3 and 6 had the highest Cohen's kappa value as these video clips were found to be the easiest to rate by some participants (Appendix 18). Some participants found that the interviewees in these video clips appeared to be very positive and appeared to know the children they were discussing very well (Appendix 18).

95% confidence intervals were also calculated to assess the difference between the ratings assigned. Ideally it was hoped that there would be no more than 0.05 of a difference between the categories assigned to each video clip. When the 95% confidence intervals of all the ratings assigned to each video clip were calculated, they were found not to be statistically significant (Figure 4 and 5) (Appendix 12 and 13). This meant there was not a significant statistical difference between the ratings assigned to each video clip.

4.6. Professional and Parental Perspectives

Some participants felt that the interviewees did not really know how well the child was doing as they were hesitant in some of their responses especially the interviewee in video clip 5 (Appendix 18). There is a possibility that the results received by professionals may be slightly inaccurate about how a child is actually doing as a result of this (O'Neill *et al*., 2004). The CAP rating scale is normally used by both parents and professionals to assess the benefits that children are receiving with their CI. A limitation of this study is that parental perspectives on how each child was doing was not assessed due to problems getting ethical approval of parents discussing their children. How children are doing with their CIs as noticed by their parents may give more insight as to how they are actually doing in their daily lives (Huttunen *et al*., 2009).

Professionals assessing children may only get a slight indication of how the child is doing as they probably spend a limited amount of time with them (Huttunen *et al*., 2009). Also professionals may respond to how the child is doing in reference to one situation as opposed to various or all situations in general which the child's parents may have more opportunity to notice if they are with the child more often (Huttunen *et al*., 2009).

Parental insight can provide important information about what is important to families, the kinds of variations that exist and why, and how views may change overtime within and across families (Huttunen *et al*., 2009). So parental perspectives could have added to this study and improved it.

One participant noticed that the child discussed in video clip 4 was able to do tasks in a controlled testing environment where there was background noise (Appendix 18). Whereas on the other hand her parents stated that she was unable to manage in a more realistic noisy environment such as the zoo (Appendix 18). This has illustrated conflicting evidence and highlights the importance of rating scales and parental insight to support findings of normal listening tests in a clinical situation. Audiological clinical assessments may not truly reflect how the child is doing at home and in school (Vidas et al., 1992; Selmi, 1985; Cunningham, 1990).

However professionals' observations are very important as well as there may be some instances where parents may have very high unrealistic expectations of their child receiving a CI. This has been reported in the past and there is also the potential for parents to be biased (Weisel *et al.,* 2006). So professionals' opinions must not be completely excluded when assessing young children.

4.7. Limitations/Modifications of the CAP-II Rating Scale

The CAP- II rating scale is a very useful and viable assessment method for assessing young children with CIs. However it has many limitations and a lot of factors need to be taken into account when using this rating scale. Some of these factors include (Stacey *et al*., 2003):

- The age of the child.
- Duration of implantation.
- The child's developmental age.
- The child's language abilities.
- The child's personality (i.e. confident, shy).
- The child's emotional state.
- The age the child was implanted.

These are only a few of the factors that need to be taken into account when using the CAP-II rating scale. They all can have an effect on how well the child can do and the benefits that they gain with their CI.

These factors all need to be considered and assessment methods such as the original CAP rating scale has been criticised for not including individual data on all of the participants discussed (Beadle *et al*., 2005). However in this study additional information on each child was provided whilst an analysis was done.

There was quite a large age range (5-18 years old) of the children discussed in the video clips. However very young children, under 5 years old were not included in this study. The older children and those who had a CI for a long period of time were assigned the highest CAP-II scores although there were some variations in this.

The majority of participants found that the CAP-II rating scale is a good assessment method. However some participants feel that the telephone categories in the rating scale are not appropriate in some instances especially when assessing young children. The reason for this is because age at implantation is getting younger and more children are gaining more benefits with their CI at a younger age yet they may not be able to use the telephone (O'Neill *et al*., 2002; Apuzzo and Yoshinaga-Itano,1995; Colletti *et al*., 2005).

Perhaps it would be better to have categories 7 and 9 as optional or separate to the rest of the rating scale. Some participants feel that the last 3 categories of the CAP-II rating scale do not follow through as well as the other categories in the scale. Some people feel that category 8 should come before category 7 as some children can do category 8 but do not use the telephone due to their age, confidence or language abilities. I.e. Some 6 year old children may be happy to use the telephone as they have confidence to do so but some 6 year old children may be too shy to use the telephone.

Two other instances were assessing use of the telephone was found not to be useful was in video clip 4 and 5. The child in video clip 4 had language difficulties which meant that it was difficult for them to use the telephone. The child in Video clip 5 was also not assessed for category 9 due to her age, 5 years old. This meant it is unknown whether she can use the telephone or not and there is a possibility that she could.

Overall this study has indicated that the CAP-II rating scale is a very valid and reliable assessment method. However two children may score the same on the CAP-II rating scale but may have different abilities. One child may be able to do a category very well, for example category 7 and another child could have the same score but may not be able to do it that well. This indicates the need for clinical tests and other assessment methods to be used in conjunction with the CAP-II in order to get a very precise accurate account of how a child is actually doing.

Although, it is not sufficient to use the CAP-II on its own to base a complete assessment of a child with CIs on (Beadle *et al.,* 2005), it is a very valid tool in that it gives professionals good insight as to how the child in coping in their daily lives with their CIs and it is useful for assessing children that are too young in order for behavioural audiological assessment to be carried out (Nikolopoulos *et al*., 2005). It is also a very effective method for following-up a child's progress over a long period of time which has been found with the original CAP scale (Archbold *et al*., 1998; Beadle et al., 2005; Wu et al., 2007).

The CAP-II rating scale can be too general in some instances and may need to be more specific. Some categories, for example category 8 does not specify whether the child can follow conversation in background noise with or without lip-reading. In some of the video clips with reference to category 8, it was asked if the child could do this task with lip-reading and in other video clips it was asked if the child could do the task without lip-reading.

Perhaps it would be better to have this category split in two one asking the question with lip-reading and the other without lip-reading. If this category is asked with and without lip-reading these are two entirely different situations. Children are more likely to be able to do it if lip-reading is an option as opposed to without lip-reading. So some children in the video clips may not have been able to achieve this rating as a result of this discrepancy. There was also found to be a similar case with categories 7 and 9 sometimes the interviewer asked about use of the telephone in background noise, in video clip 3 and in other cases she didn't.

However it is not stated in the CAP-II rating scale whether or not it is use of the telephone in background noise. So these few categories of the CAP-II rating scale may need to be made more specific in order to get a true accurate account of the benefits that children are receiving with their CIs.

4.8. Future Research

It is suggested that in the future test-retest reliability should be conducted on the modified CAP-II rating scale with a wider participant population and also including more case children discussed. It may also be useful to include younger children, those under the age of 5 years in future research to assess whether or not children under this age group are capable of achieve high CAP-II scores. The reason for this is because more and more children are being fitted with CIs at a younger age. The CAP-II rating scale is designed to assess very young children. Children under 5 years of age have not been considered in this study however.

Future studies of the CAP-II rating scale should also assess young children over a long period of time to consider whether or not the CAP-II rating scale is a valid measure for assessing young children beyond 5 years up until 10 years and perhaps longer.

Investigation of both parental and professional perspectives of children should be conducted to assess if there is a large difference between the ratings assigned by parents and professionals. More information and more accurate ratings may be assigned to each child based on their parents' perspectives if they are included in the study.

In future studies the questions in the interview need to be asked in a more controlled way. More measures need to be taken to reduce influencing responses from interviewees which may cause bias and result in inaccurate accounts of how the child is actually doing.

As it is also unknown the extent to which binaural cues are available to listeners and used effectively, which indicates that the knowledge in this area is very limited. Future work also needs to address the extent to which synchronisation and enhanced binaural hearing are functionally useful to those using bilateral CIs (Litovsky *et al.,* 2004b). So that the CAP-II rating scale can account for all these benefits if it is used to assess children with bilateral CIs.

5. Conclusion

- This study shows that the CAP-II rating scale is reliable as shown by the high Cohen's kappa values and correlation coefficients which indicate high inter-rater reliability.

- It also shows that the extended CAP, the CAP-II provides a reliable and valid measure of the benefits that children are now able to achieve with their CIs and bilateral fittings. It gives a good indication of how well children are now able to progress and do well with the CIs.

- The results indicated that there was no ceiling effects occurring as all of the video clips were not rated the highest category, category 9 by all participants included in the study. Even though 2 video clips were assigned category 9 by the majority of participants some participants give a lower CAP-II rating.

- The two additional categories, categories 8 and 9 have proven to be very valuable in assessing the additional benefits that have now become apparent with recent CIs and bilateral fittings. However they may need to be made more specific as it is unclear whether category 9 is referring to using the telephone in background noise or in quiet. In some video clips the interviewer asked could the child do the task in background noise however in other video clips she asked could the child do it in quiet. This is too different situations and it is unclear from the rating scale whether it is referring to using the telephone in quiet or in noise.

- Category 8 should come before category 7 as most young children are too young to use the telephone or are unable to due to other factors such as speech difficulties. So therefore they are unable to fulfil category 7 yet they may be able to do category 8 which is listening in a reverberant room which has been shown in video clip 5. As a result of this the child may be given a rating much lower than their actual ability because the CAP-II rating scale is a hierarchy scale, meaning that if a child cannot do level 7 they may only be given a rating of 6 even though they are capable of doing category 8. So the last 3 categories may be better if they were reordered in order to account for this.

- When the CAP-II rating scale is used parents and professionals that are using it must also consider all the factors, discussed in the discussion that may have an effect on the ratings assigned.

- Also with the changes outlined throughout the discussion the CAP-II could prove to be an even more reliable assessment method put into clinical use to assess the benefits of young children over a prolonged period of time.

References

Allen, C., Nikolopoulos, T. P., Dyar, D, O'Donoghue, G. M. (2001). Reliability of a rating scale for measuring speech intelligibility after paediatric cochlear implantation. *Otology and Neurotology*. 22: 631-633.

Answers.com. (2010). Floor Effects. In *Sports Science and Medicine.* Available online: http://www.answers.com/topic/floor-effects [18/05/2010].

Apuzzo, M., and Yoshinaga-Itano, C. (1995). Early indentification of infants with significant hearing loss and the Minnesota child development inventory. *Semin Hear*. 16: 124-39.

Archbold, S., Lutman, M., Marshall, D. (1995). Categories of auditory performance. *Ann Otol Rhinol Laryngol.* 104 (suppl 166): 312-4.

Archbold, S., Lutman, M. E., Nikolopoulos, T. P. (1998). Categories of auditory performance: inter-user reliability. *British Journal of Audiology.* 32: 7-12.

Archbold, S., O'Donoghue, G. M. (2009). Cochlear implantation in children: current status. *Paediatrics and Child Health.* 19 (10): 457-463.

Archbold, S., Robinson, K. (1997). A European perspective on paediatric cochlear implantation, rehabilitation services, and their educational implications. *The Americain Journal of Otology.* 18: S75-78.

Beadle, E. A. R., McKinley, D, J., Nikolopoulos, T. P., Brough, J., O'Donoghue, G. M., Archbold, S. (2005). Long-term functional outcomes and academic-occupational status of implanted children after 10 to 14 years of cochlear implant use. *Otology & Neurotology*. 26: 1152-1160.

Brennan, R. L., and Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement.* 41: 687-699.

Cnet Reviews. (2010). Cnet archive. *Creative SBS 50- Speakers.* Available online: http://reviews.cnet.com/pc-speakers/creative-sbs-50-speaker/1707-3179_7-30209526.html [25/08/10].

Colletti, L. (2009). Long-term follow-up of infants (4-11 months) fitted with cochlear implants. *Act Oto-Laryngologica.* 129: 361-366.

Colletti, V., Carner, M., Miorelli, V., Guida, M., Colletti, L., Fiorino, F. G. (2005). Cochlear implantation at under 12 months: report on 10 patients. *Laryngoscope*. 115: 445-449.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement.* 20: 37-46.

Cordes, A. K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research.* 37: 264-278.

Cunningham, J. K. (1990). Parent's evaluation of the effects of 3M/House cochlear implantation at under 12 months: report on 10 patients. *Laryngoscope.* 115: 445-449.

Davis, A. Parving, A. (1994). Towards appropriate epidemiology data on childhood hearing disability: a comparative European study of birth cohorts 1982–1988. *J Audiol Med* 3: 35–47.

Encyclopaedia Britannica (2010). Deaf-mutism. In *Encyclopaedia Britannica*. Available online: http://www.britannica.com/EBchecked/topic/154320/deaf-mutism [06/05/2010].

Govaerts, P.J., Beukelaer, C. D., Daemers, K., Ceulaer, G. D., Yperman, M., Somers, T., Schatteman, I., Offeciers, F. E. (2002). Outcome of cochlear implantation at different ages from 0 to 6 years. *Otology & Neurotology*. 23: 885-890.

Huttunen, K., Rimmanen, S., Vikman, S., Virokannas, N., Sorri, M., Archbold, S., Lutman, M. E. (2009). Parents' views on the quality of life of their children 2-3 years after cochlear implantation. *International Journal of Pediatric Otorhinolaryngology*. 73: 1786-1794.

Hyde, M. L. (2000). Reasonable psychometric standards for self-report outcome measures in Audiological rehabilitation. *Ear & Hearing.* 21: 24S-36S.

Inscoe, J. (1999). Communication outcomes after paediatric cochlear implantation. *International Journal of Pediatric Otorhinolaryngology*. 47: 195-200.

Kim, L-S., Jang, Y. S., Choi, A-H., AHN, S-Y., Park, J-S., Lee, Y-M., Jeong,S-W. (2009). Bilateral cochlear implants in children. *Cochlear Implants International.* 10 (S1): 74-77.

Kinner, P. R., and Gray, C. D. (2004). SPSS 12. Made simple. Psychology Press Taylor and Francis group. Hove and New York. 185-189.

Kishon-Rabin, L., Taitelbaum-Swead, R., Ezrati-Vinacour, R., Hildesheimer, M. (2005). Prelexical vocalization in normal hearing and hearing-impaired infants before and after cochlear implantation and it relation to early auditory skills. *Ear and Hearing.* 26: 17S-29S.

Kolmogorov–Smirnov goodness-of-fit test. (2003-2010). *In Engineering Statistics Handbook*. Available online: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm [25/08/2010].

Kuhn-Inacker, H., Shehata-Dieler, W., Muller, J., Helms, L. (2004). Bilateral cochlear implants: A way to optimize auditory perception abilities in deaf children. *International Journal of Pediatric Otorhinolaryngology*. 68 (10): 1257 - 1266.

Litovsky, R., Johnstone, P., Parkinson, A., Peters, R., Lake, J. (2004a). Bilateral cochlear implants in children. *International congress series*. 1273: 451-454.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., Yu, G. (2004b). Bilateral cochlear implants in adults and children. *Arch Otolaryngol Head Neck Surg*. 130: 648-655.

Mauk, G. W., Behrens, T. R. (1993). Historical, political, and technological context associated with early identification of hearing loss. *Semin Hear.* 14: 1–17.

Miyamoto, R. T., Svirsky, M. A., Robbins, A. M. (1997). Enhancement of expressive language in prelingually deaf children with cochlear implants. *Acta otolaryngol. 117*: 154-7.

Nicholas, J. (1994). Sensory aid use and the development of communicative function. *The Volta Review.* 96: 181-198.

Nicholas, J., & Geers, A. (1997). Communcation of oral deaf and normally-hearing children at 36 months of age. *Journal of Speech, Language, and Hearing Research*. 40: 1314-1327.

Nicholas, J., Geers, A., Kozak, V. (1994). Development of communicative function in young hearing-impaired and normally-impaired children. *The Volta Review*. 96: 113-135.

Nicholas, J. G, Geers, A. E., Rollins, P. R. (1999). Inter-rater reliability as a reflection of ambiguity in the communication of deaf and normally-hearing children. *J. Commun. Disord*. 32: 121-134.

National Institute on Deafness and Other Communication Disorders (NIDCD) (2008). *Better Hearing in Real-Life Situations: Advances in Cochlear Implants.* Available online: http://www.nidcd.nih.gov/research/stories/archives/06/09_06_07.htm [18/05/2010].

National Institute for Health and Clinical Excellence (NICE) (2009). *Hearing impairment – cochlear implants: Full guidance*. Cochlear implants for children and adults with severe to profound deafness. Available online: http://www.nice.org.uk/TA166 [14/03/2010].

Nicholas, J. G., Geers, A. E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *J. Speech Lang Hear Res.* 50: 1048-62.

Nikolopoulos, T. P., Archbold, S. M., Gregory, S. (2005). Young deaf children with hearing aids or cochlear implants: early assessment package for monitoring progress*. International Journal of Paediatric Otorhinolaryngology*. 69: 175-186.

Nikolopoulos, T. P., O'Donoghue, G. M., Archbold, S. (1999a). Age at implantation: its importance in paediatric cochlear implantation. *Laryngoscope.* 109: 595-599.

Nikolopoulos, T. P., O'Donoghue, G. M., Archbold, S. (1999b). The development of auditory perception in children following cochlear implantation. *International Journal of Paediatric Otorhinolaryngology.* 49 (Suppl. 1): S189-S191.

Oller, D. K., Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development.* 59: 441-449.

Oller, D. K., Eilers, R. E., Bull, D. H., and Carney, A. E. (1985). Prespeech vocalisations of a deaf infant: a comparison with normal methaphonological development. *Journal of Speech, Language, and Hearing Research*. 28: 47-63.

O'Neill, C., Lutman, M. E., Archbold, S. M., Gregory, S., Nikolopoulos, T. P. (2004). Parents and their cochlear implanted child: questionnaire development to assess parental views and experiences. *International Journal of Pediatric Otorhinolaryngology*. 68: 149-160.

O'Neill, C., O'Donoghue, G. M., Archbold, S. M., Nikolopoulos, T. P., Sach, T. (2002). Variations in gains in auditory performance from paediatric cochlear implantation. *Otology & Neurotology.* 23: 44-48.

Parving, A. (1993). Congenital hearing disability: epidemiology and identification. A comparison between two health authority districts. *Int J Pediatr Otorhinolaryngol* 27: 29–46.

Price runner (1999-2010). Product information: *Sony DCR-SR55*. Available online: www.pricerunner.co.uk/pi/8-1059779/camcorders/Sony-DCR-SR55-Product-Info [25/08/2010].

Randolph, J. J. (2008). *Online Kappa Calculator.* Available online: http://www.justusrandolph.net/kappa/ [25/08/2010].

Rowans University (2010). Confidence intervals, error bars and sample size for a single mean. Available online : http://users.rowan.edu/~everett/courses/Sys_3/docs/11%20Conf%20Int%20&%20Sam%20Size.pdf [06/09/2010].

Schauwers, k., Gillis, S., Daemers, K., Beukelaer, C. D., Govaerts, P. (2004a). Cochlear implantation between 5 and 20 months of age: the onset of babbling and the audiologic outcome. *Otology & Neurotology*. 25: 263-270.

Schauwers, k., Gillis, S., Daemers, K., Beukelaer, C. D., Ceulaer, G. D., Yperman, M., Govaerts, P. J. (2004b). Normal hearing and language development in a deaf-born child. *Otology & Neurotology*. 25: 924-929.

Selmi, A. (1985). Monitoring and evaluating the educational effects of the cochlear implant. *Ear and Hear.* 6 (3): 52S-59S.

Sharma, A., Dorman, M., Spahr, T. (2002c). A sensitive period for the development of the central auditory system in childrenwith cochlear implants. *Ear Hear.* 23 (6): 532-539.
Spearman's Rank Correlation Coefficient (2009). In *Barcelona Field Studies Centre.* Available online: http://geographyfieldwork.com/SpearmansRank.htm [25/08/10].

Stacey, P. C., Fortnum, H. M., Barton, G. R., Summerfield, A. Q. (2003). Use of ordinal regression to assess the influence of cochlear implantation on the categories of auditory performance. Cochlear Implants International. 4 (Suppl. 1): 64-65.

Stein, L. K. (1999). Factors influencing the efficacy of universal newborn hearing screening. *Pediatr Clin North Am* 46: 95–105.

Taitelbaum-Sweed, R., Kishon-Rabin, L., Kaplan-Neeman, R., Muchnik, C., Kronenberg, J., Hildesheimer, M. (2005). Speech perception of children using Nucleus, Clarion or Med-El cochlear implants. *Pediatric Otorhinolaryngology*. 69: 1675-1683.

The Nemours Foundation (1995-2010). KidsHealth from Nemours – for parents – cochlear implants. Available online: http://kidshealth.org/parent/general/eyes/cochlear.html [17/05/2010].

Vidas, S., Hassan, R., Parnes, L. S. (1992). Real-life performance considerations of four paediatric multi-channel cochlear implant recipients. *J. Otol*. 21: 387-93.

Weisel, A., Most, T., Micheal, R. (2006). Mother's stress and expectations as a function of time since child's cochlear implantation. *J. Deaf Stud. Deaf Educ.* 12 (1): 55-64.

Wikia Education (2010). Ceiling effects. In *The psychology Wiki*. Available online: http://psychology.wikia.com/wiki/Ceiling_effect [18/05/2010].

Windows 7 Features (2010). In *Windows media player 12.* Available online: http://windows.microsoft.com/en-US/windows7/products/features/windows-media-player-12 [25/08/10].

Wu, C. –M., Sun, Y. –S., Liu, T. –C (2007). Long-term categorical auditory performance and speech intelligibility in Mandarin-speaking prelingually deaf children with early cochlear implantation in Taiwan. *Clinical Otolaryngology*. 33: 32-55.

Appendix 1:
DVD of the 6 Video Clips

Appendix 2:

Introductory Video Clip Script

- Hi, this is just an introductory video.
- Following this introduction you will be shown 6 video clips lasting 3-5 minutes each.
- Before we begin make yourself comfortable and adjust your seat and screen to a position that is most suitable for you.
- I would also like you to adjust the volume to a comfortable level.
- Before watching the following videos please read through the instructions that you are given.
- When you are comfortable and happy with what you have to do, you will be shown the videos clips.

Appendix 3:
INSTRUCTIONS

The Categories of Auditory Performance (CAP)-II is a rating scale used to assess the benefits that children receive with use of cochlear implants in everyday situations.

PLEASE READ THROUGH THE FOLLOWING INSTRUCTIONS BELOW BEFORE BEGINNING.

1. I am going to show 6 video clips each 3-5 mins long.

2. The video clips show different professionals talking about 6 different children and how they are getting on with their cochlear implants.

3. You will be given the CAP-II rating scale. (Shown below).

PLEASE READ THROUGH THIS TABLE THOROUGHLY AND TICK <u>ONE</u> OF THE FOLLOWING.

Category	Criteria	Tick ONE
0	No awareness of environmental sounds.	
1	Awareness of environmental sounds.	
2	Response to speech sounds (e.g. ''go'').	
3	Identification of environmental sounds.	
4	Discrimination of some speech sounds without lip-reading.	
5	Understanding of common phrases without lip-reading.	
6	Understanding of conversation without lip-reading.	
7	Use of telephone with known listener.	
8	*Follows group conversation in a reverberant room or where there is some interfering noise, such as a classroom or restaurant.*	
9	*Use of phone with unknown speaker in unpredictable context.*	

4. You are required to tick **<u>ONE</u>** of the categories that you think applies most to each individual child discussed in the video clips.

5. Read through the CAP-II rating scale and if there is any terminology you do not understand, ask me.

THANK YOU!

Appendix 4:

Participant Record Sheet

Participant Name/ No: ________________ **Code** (Initials)**:** _________

First Language: _________________ **Age:** _______

Date: ________

Video Clip Number: _________

PLEASE READ THROUGH THIS TABLE THOROUGHLY AND TICK ONE OF THE FOLLOWING.

Category	Criteria	Tick ONE
0	No awareness of environmental sounds.	
1	Awareness of environmental sounds.	
2	Response to speech sounds (e.g. ‘‘go’’).	
3	Identification of environmental sounds.	
4	Discrimination of some speech sounds without lip-reading.	
5	Understanding of common phrases without lip-reading.	
6	Understanding of conversation without lip-reading.	
7	Use of telephone with known listener.	
8	*Follows group conversation in a reverberant room or where there is some interfering noise, such as a classroom or restaurant.*	
9	*Use of phone with unknown speaker in unpredictable context.*	

Please write any additional comments in the box below.

Subject Comments:

Investigators Comments:

Investigators Record Sheet

Subject Name/No: ____________

First Language: _________________

Date: ________

Code (Initials)**:** _________

Male/Female: _________

Age: _______

Start Time: _____________

Finish Time: ____________

Number of Repetitions Required:

Video Clip No:	1	2	3	4	5	6
No of Repeats:						

Investigators Comments:

Appendix 6:

Risk Assessment & Safety and Ethics Approval

UNIVERSITY OF Southampton
Institute of Sound and Vibration Research

RISK ASSESSMENT

FORM - A

This Form Does Not Permit
Late Working (after 18:00 hrs)

Record File No. 10/183

Originator - Louise Gilmour E-mail lag1g09@soton.ac.uk Ext. ____________

Signature - L. Gilmour

Title and location of procedure

Inter-rater reliability of categories of Auditory Performance (CAP)-II

Nature of Risk

	Nature of Risk	Risk rating score $L \times S$ Please see below
1)	Problems with the laptop that will be used to show the videos.	1
2)		
3)		
4)		

Control Measures for above Risks

	Control Measures	Risk rating score $L \times S$ Please see below
1)	Laptop will have to undergo a health and safety check.	1
2)		
3)		
4)		

Description and Rig Operating Procedures

Please attach (on a separate sheet) a description of your activity and any rig operating procedures as appropriate

Named Authorised Operators

LOUISE GILMOUR

Any Hazardous Materials in use? NO   Is COSHH assessment attached? N/A

Is Safety & Ethics approval required? YES

Supervisor's Signature	Carl Versluis	Date	02/06/10
Print	Carl Verschuur		
Safety Advisor's Approval	N. Davis	Date	26-6-10
Head of School's Signature	pp [illegible]	Date	9-7-10
Expiry / review Date	June 11		

UNIVERSITY OF SOUTHAMPTON

INSTITUTE OF SOUND AND VIBRATION RESEARCH

HUMAN EXPERIMENTATION SAFETY AND ETHICS COMMITTEE

Human Experimentation Safety and Ethics Application Number: 1110

Title of Experiment: Inter-rater reliability of Categories of Audiotry Performance (CAP)-II

submitted by: Louise Gilmour **on** 3rd June 2010

The Human Experimentation Safety and Ethics Committee has found the planned study satisfactory and confirm that it can proceed.

Please observe the following:

1. ***Record the Human Experimentation Safety and Ethics Approval Number on the subject consent forms. This number can be found at the top of this page.***
2. ***The subject consent forms, to be completed by all the subjects who participate in this study, must be provided to the Secretary of the Human Experimentation Safety and Ethics Committee by 25th September 2010.***
3. ***You must not make any changes to the study without obtaining the approval of the Human Experimentation Safety and Ethics Committee.***
4. ***You must inform the Human Experimentation Safety and Ethics Committee immediately if any subject experiences a problem or makes a complaint related to this study.***

12 July

Date: ~~25th June~~ 2010

Signed: ....

Professor M.J. Griffin
Chair, Human Experimentation Safety and Ethics Committee

cc: Supervisor: Julie Brinton, Carl Verschuur

Appendix 7:

Order that Video Clips Were Viewed by Each Participant

Subject Number	Video Order					
1	1	2	6	3	5	4
2	2	3	1	4	6	5
3	3	4	2	5	1	6
4	4	5	3	6	2	1
5	5	6	4	1	3	2
6	6	1	5	2	3	4
7	1	3	4	5	2	6
8	2	1	3	6	5	4
9	3	5	6	1	2	4
10	4	6	5	1	3	2
11	5	2	1	6	4	3
12	6	4	3	5	1	2
13	1	4	2	6	3	5
14	2	5	4	1	3	6
15	3	6	1	2	4	5
16	4	2	6	5	1	3
17	5	4	2	3	1	6
18	6	3	1	4	2	5
19	1	5	3	4	6	2
20	2	6	5	1	4	3
21	3	1	4	6	5	2
22	4	1	2	3	5	6
23	5	1	3	2	6	4
24	6	2	4	1	5	3
25	1	6	5	2	4	3
26	2	4	6	5	3	1
27	3	2	5	4	6	1
28	4	3	1	2	6	5
29	5	3	6	4	2	1
30	6	5	2	3	4	1
31	1	2	4	6	5	3
32	2	3	6	4	1	5
33	3	4	1	2	6	5

Appendix 8:
Participant Information

Participant No.	Duration (Mins)	Age (Years)	Gender
1	22	22	Female
2	36	25	Male
3	25	19	Female
4	30	24	Male
5	21	19	Female
6	22	25	Male
7	24	47	Female
8	23	22	Female
9	28	25	Male
10	27	22	Female
11	22	29	Female
12	20	23	Female
13	21	25	Female
14	24	22	Female
15	41	22	Female
16	20	23	Female
17	27	25	Female
18	29	24	Female
19	25	50	Male
20	22	20	Female
21	42	33	Male
22	27	50	Female
23	20	74	Female
24	20	19	Female
25	19	34	Female
26	22	33	Female
27	22	24	Male
28	27	50	Female
29	22	33	Male
30	23	24	Female
31	21	26	Female
32	21	23	Female
33	21	23	Female
Average	24.73	29.06	
Gender of Participants		**Male**	**Female**
Number		8	25

Appendix 9:

Distribution of Data

One-Sample Kolmogorov-Smirnov Test

		Video 1	Video 2	Video 3	Video 4	Video 5	Video 6
N		33	33	33	33	33	33
Normal Parameters[a,,b]	Mean	6.6970	8.9394	7.9091	6.6970	7.6667	8.9091
	Std. Deviation	.72822	.24231	.38435	.52944	.85391	.29194
Most Extreme Differences	Absolute	.510	.538	.533	.444	.470	.531
	Positive	.339	.401	.407	.284	.348	.378
	Negative	-.510	-.538	-.533	-.444	-.470	-.531
Kolmogorov-Smirnov Z		2.929	3.091	3.061	2.549	2.700	3.052
Asymp. Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
Exact Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
Point Probability		.000	.000	.000	.000	.000	.000

a. Test distribution is Normal.

b. Calculated from data.

Appendix 10:

Number of Repetitions

Video	1	2	3	4	5	6
Participant No						
1	0	0	0	0	0	0
2	1	0	1	0	1	0
3	0	0	0	0	0	0
4	0	0	1	0	0	0
5	0	0	0	0	0	0
6	1	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	1	1	1	1	1	1
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0
20	0	0	0	0	0	0
21	0	0	0	1	0	0
22	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	0	0	0
26	0	0	0	0	0	0
27	0	0	0	0	0	0
28	0	0	0	0	0	0
29	0	0	0	0	0	0
30	0	0	0	0	0	0
31	0	0	0	0	0	0
32	0	0	0	0	0	0
33	0	0	0	0	0	0
Average No. Repeats	0.09	0.03	0.09	0.06	0.06	0.03

Appendix 11:

Participant Information Native English Participants

Participant No.	Duration (Mins)	Age (Years)	Gender
1	22	22	Female
2	25	19	Female
3	30	24	Male
4	21	19	Female
5	24	47	Female
6	23	22	Female
7	27	22	Female
8	22	29	Female
9	21	25	Female
10	24	22	Female
11	41	22	Female
12	20	23	Female
13	29	24	Female
14	25	50	Male
15	22	20	Female
16	42	33	Male
17	27	50	Female
18	20	74	Female
19	20	19	Female
20	22	24	Male
21	27	50	Female
22	23	24	Female
23	21	26	Female
24	21	23	Female
25	21	23	Female
Average	24.8	29.44	

Appendix 12:

Descriptive Statistics

			Statistic	Std. Error
video1	Mean		6.6970	.12677
	95% Confidence Interval for Mean	Lower Bound	6.4388	
		Upper Bound	6.9552	
	5% Trimmed Mean		6.7744	
	Median		7.0000	
	Variance		.530	
	Std. Deviation		.72822	
	Minimum		5.00	
	Maximum		7.00	
	Range		2.00	
	Interquartile Range		.00	
	Skewness		-2.038	.409
	Kurtosis		2.287	.798
video2	Mean		8.9394	.04218
	95% Confidence Interval for Mean	Lower Bound	8.8535	
		Upper Bound	9.0253	
	5% Trimmed Mean		8.9882	
	Median		9.0000	
	Variance		.059	
	Std. Deviation		.24231	
	Minimum		8.00	
	Maximum		9.00	
	Range		1.00	
	Interquartile Range		.00	
	Skewness		-3.861	.409
	Kurtosis		13.736	.798
Video3	Mean		7.9091	.06691
	95% Confidence Interval for Mean	Lower Bound	7.7728	
		Upper Bound	8.0454	

	5% Trimmed Mean		7.9882	
	Median		8.0000	
	Variance		.148	
	Std. Deviation		.38435	
	Minimum		6.00	
	Maximum		8.00	
	Range		2.00	
	Interquartile Range		.00	
	Skewness		-4.503	.409
	Kurtosis		20.828	.798
Video4	Mean		6.6970	.09216
	95% Confidence Interval for Mean	Lower Bound	6.5092	
		Upper Bound	6.8847	
	5% Trimmed Mean		6.7525	
	Median		7.0000	
	Variance		.280	
	Std. Deviation		.52944	
	Minimum		5.00	
	Maximum		7.00	
	Range		2.00	
	Interquartile Range		1.00	
	Skewness		-1.553	.409
	Kurtosis		1.679	.798
Video5	Mean		7.6667	.14865
	95% Confidence Interval for Mean	Lower Bound	7.3639	
		Upper Bound	7.9695	
	5% Trimmed Mean		7.8081	
	Median		8.0000	
	Variance		.729	
	Std. Deviation		.85391	
	Minimum		4.00	
	Maximum		8.00	
	Range		4.00	
	Interquartile Range		.00	

	Skewness		-3.123	.409
	Kurtosis		10.674	.798
Video6	Mean		8.9091	.05082
	95% Confidence Interval for Mean	Lower Bound	8.8056	
		Upper Bound	9.0126	
	5% Trimmed Mean		8.9545	
	Median		9.0000	
	Variance		.085	
	Std. Deviation		.29194	
	Minimum		8.00	
	Maximum		9.00	
	Range		1.00	
	Interquartile Range		.00	
	Skewness		-2.983	.409
	Kurtosis		7.343	.798

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
video1	33	6.6970	.72822	5.00	7.00
video2	33	8.9394	.24231	8.00	9.00
Video3	33	7.9091	.38435	6.00	8.00
Video4	33	6.6970	.52944	5.00	7.00
Video5	33	7.6667	.85391	4.00	8.00
Video6	33	8.9091	.29194	8.00	9.00

Appendix 13:

Descriptive Statistics Native English Participants

			Statistic	Std. Error
Video1	Mean		6.7600	.13266
	95% Confidence Interval for Mean	Lower Bound	6.4862	
		Upper Bound	7.0338	
	5% Trimmed Mean		6.8444	
	Median		7.0000	
	Variance		.440	
	Std. Deviation		.66332	
	Minimum		5.00	
	Maximum		7.00	
	Range		2.00	
	Interquartile Range		.00	
	Skewness		-2.491	.464
	Kurtosis		4.563	.902
Video2	Mean		8.9600	.04000
	95% Confidence Interval for Mean	Lower Bound	8.8774	
		Upper Bound	9.0426	
	5% Trimmed Mean		9.0000	
	Median		9.0000	
	Variance		.040	
	Std. Deviation		.20000	
	Minimum		8.00	
	Maximum		9.00	
	Range		1.00	
	Interquartile Range		.00	
	Skewness		-5.000	.464
	Kurtosis		25.000	.902
Video3	Mean		7.9200	.08000
	95% Confidence	Lower Bound	7.7549	

	Interval for Mean	Upper Bound	8.0851	
	5% Trimmed Mean		8.0000	
	Median		8.0000	
	Variance		.160	
	Std. Deviation		.40000	
	Minimum		6.00	
	Maximum		8.00	
	Range		2.00	
	Interquartile Range		.00	
	Skewness		-5.000	.464
	Kurtosis		25.000	.902
Video4	Mean		6.7200	.09165
	95% Confidence Interval for Mean	Lower Bound	6.5308	
		Upper Bound	6.9092	
	5% Trimmed Mean		6.7444	
	Median		7.0000	
	Variance		.210	
	Std. Deviation		.45826	
	Minimum		6.00	
	Maximum		7.00	
	Range		1.00	
	Interquartile Range		1.00	
	Skewness		-1.044	.464
	Kurtosis		-.998	.902
Video5	Mean		7.5600	.19218
	95% Confidence Interval for Mean	Lower Bound	7.1634	
		Upper Bound	7.9566	
	5% Trimmed Mean		7.7111	
	Median		8.0000	
	Variance		.923	
	Std. Deviation		.96090	
	Minimum		4.00	
	Maximum		8.00	
	Range		4.00	

	Interquartile Range		.50	
	Skewness		-2.635	.464
	Kurtosis		7.476	.902
Video6	Mean		8.9600	.04000
	95% Confidence Interval for Mean	Lower Bound	8.8774	
		Upper Bound	9.0426	
	5% Trimmed Mean		9.0000	
	Median		9.0000	
	Variance		.040	
	Std. Deviation		.20000	
	Minimum		8.00	
	Maximum		9.00	
	Range		1.00	
	Interquartile Range		.00	
	Skewness		-5.000	.464
	Kurtosis		25.000	.902

Appendix 14: Ratings Assigned

	Video No.					
Participant No.	**1**	**2**	**3**	**4**	**5**	**6**
1	7	9	8	6	8	9
2	7	8	7	6	8	9
3	7	9	8	7	8	9
4	7	9	8	7	8	9
5	7	9	8	7	8	9
6	5	9	8	7	8	8
7	7	9	8	6	8	9
8	7	9	8	7	8	9
9	7	9	8	7	8	9
10	7	9	8	7	6	9
11	7	9	8	6	8	9
12	7	9	8	7	8	9
13	5	9	8	7	6	9
14	7	9	8	7	8	9
15	7	9	8	7	8	9
16	7	9	8	6	8	9
17	7	9	8	7	8	9
18	7	9	8	7	8	9
19	7	8	8	7	7	9
20	5	9	8	7	8	9
21	5	9	8	7	8	9
22	7	9	8	6	7	9
23	7	9	8	6	8	9
24	7	9	8	7	8	9
25	7	9	8	7	8	9
26	5	9	8	5	8	8
27	7	9	8	7	8	9
28	7	9	8	7	8	9
29	7	9	8	7	8	9
30	7	9	6	7	7	9
31	7	9	8	7	8	9
32	7	9	8	6	4	8
33	7	9	8	7	8	9
Average CAP-II Score	6.697	8.939	7.909	6.697	7.667	8.909
Standard Error						
Standard Deviation	0.728	0.242	0.384	0.529	0.854	0.292
Confidence Interval (+/-)	0.248	0.083	0.131	0.181	0.291	0.100

Appendix 15:

Ratings Assigned Native English Participants

	Video No.					
Participant No.	**1**	**2**	**3**	**4**	**5**	**6**
1	7	9	8	6	8	9
2	7	9	8	7	8	9
3	7	9	8	7	8	9
4	7	9	8	7	8	9
5	7	9	8	6	8	9
6	7	9	8	7	8	9
7	7	9	8	7	6	9
8	7	9	8	6	8	9
9	5	9	8	7	6	9
10	7	9	7	8	8	9
11	7	9	8	7	8	9
12	7	9	8	6	8	9
13	7	9	8	7	8	9
14	7	8	8	7	7	9
15	5	9	8	7	8	9
16	5	9	8	7	8	9
17	7	9	8	6	7	9
18	7	9	8	6	8	6
19	7	9	8	7	8	9
20	7	9	8	7	8	9
21	7	9	8	7	8	9
22	7	9	6	7	7	9
23	7	9	8	7	8	9
24	7	9	8	6	4	8
25	7	9	8	7	8	9
Average CAP-II Score	6.54	8.69	7.69	6.65	7.46	8.73
Standard Deviation	0.66	0.20	0.44	0.52	0.96	0.62
Confidence Interval (+/-)	0.26	0.08	0.17	0.20	0.38	0.24

Appendix 16:
Correlation Coefficient

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	Average
1	Correlation Coefficient	1.000	.894*	.985**	.985**	.985**	.844*	1.000**	.985**	.985**	.727	1.000**	.985**	.776	.677	.985**	1.000**	.985**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.687	.985**	.702
	Sig. (2-tailed)	.	.016	.000	.000	.000	.035	.	.000	.000	.101	.	.000	.070	.140	.000	.	.000	.000	.043	.005	.005	.003	.421	.000	.000	.018	.000	.000	.000	.175	.000	.132	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
2	Correlation Coefficient	.894*	1.000	.862*	.862*	.862*	.625	.894*	.862*	.862*	.455	.894*	.862*	.522	.677	.862*	.894*	.862*	.862*	.636	.803	.803	.773	.182	.862*	.862*	.683	.862*	.862*	.862*	.683	.862*	.343	.862*	.630
	Sig. (2-tailed)	.016	.	.027	.027	.027	.184	.016	.027	.027	.365	.016	.027	.288	.140	.027	.016	.027	.027	.175	.054	.054	.072	.730	.027	.027	.135	.027	.027	.027	.135	.027	.505	.027	.104
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
3	Correlation Coefficient	.985**	.862*	1.000	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
4	Correlation Coefficient	.985**	.862*	1.000**	1.000	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
5	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
6	Correlation Coefficient	.844*	.625	.889*	.889*	.889*	1.000	.844*	.889*	.889*	.563	.844*	.889*	.770	.635	.889*	.844*	.889*	.889*	.623	.907*	.907*	.750	.657	.889*	.889*	.984**	.889*	.889*	.889*	.426	.889*	.585	.889*	.663
	Sig. (2-tailed)	.035	.184	.018	.018	.018	.	.035	.018	.018	.245	.035	.018	.073	.176	.018	.035	.018	.018	.186	.013	.013	.086	.157	.018	.018	.000	.018	.018	.018	.399	.018	.222	.018	.068
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
7	Correlation Coefficient	1.000**	.894*	.985**	.985**	.985**	.844*	1.000	.985**	.985**	.727	1.000**	.985**	.776	.677	.985**	1.000**	.985**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.687	.985**	.702
	Sig. (2-tailed)	.	.016	.000	.000	.000	.035	.	.000	.000	.101	.	.000	.070	.140	.000	.	.000	.000	.043	.005	.005	.003	.421	.000	.000	.018	.000	.000	.000	.175	.000	.132	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
8	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
9	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

10	Correlation Coefficient	.727	.455	.739	.739	.739	.563	.727	.739	.739	1.000	.727	.739	.896*	.554	.739	.727	.739	.739	.906*	.727	.727	.864*	.091	.739	.739	.572	.739	.739	.739	.572	.739	.940**	.739	.689
	Sig. (2-tailed)	.101	.365	.094	.094	.094	.245	.101	.094	.094	.	.101	.094	.016	.254	.094	.101	.094	.094	.013	.101	.101	.027	.864	.094	.094	.235	.094	.094	.094	.235	.094	.005	.094	.136
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
11	Correlation Coefficient	1.000**	.894*	.985**	.985**	.985**	.844*	1.000**	.985**	.985**	.727	1.000	.985**	.776	.677	.985**	1.000**	.985**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.687	.985**	.702
	Sig. (2-tailed)	.	.016	.000	.000	.000	.035	.	.000	.000	.101	.	.000	.070	.140	.000	.	.000	.000	.043	.005	.005	.003	.421	.000	.000	.018	.000	.000	.000	.175	.000	.132	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
12	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
13	Correlation Coefficient	.776	.522	.849*	.849*	.849*	.770	.776	.849*	.849*	.896*	.776	.849*	1.000	.728	.849*	.776	.849*	.849*	.892*	.896*	.896*	.806	.134	.849*	.849*	.720	.849*	.849*	.849*	.564	.849*	.779	.849*	.702
	Sig. (2-tailed)	.070	.288	.033	.033	.033	.073	.070	.033	.033	.016	.070	.033	.	.101	.033	.070	.033	.033	.017	.016	.016	.053	.800	.033	.033	.107	.033	.033	.033	.244	.033	.068	.033	.081
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
14	Correlation Coefficient	.677	.677	.750	.750	.750	.635	.677	.750	.750	.554	.677	.750	.728	1.000	.750	.677	.750	.750	.516	.800	.800	.554	.000	.750	.750	.581	.750	.750	.750	.904*	.750	.364	.750	.685
	Sig. (2-tailed)	.140	.140	.086	.086	.086	.176	.140	.086	.086	.254	.140	.086	.101	.	.086	.140	.086	.086	.294	.056	.056	.254	1.000	.086	.086	.227	.086	.086	.086	.013	.086	.478	.086	.156
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
15	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
16	Correlation Coefficient	1.000**	.894*	.985**	.985**	.985**	.844*	1.000**	.985**	.985**	.727	1.000**	.985**	.776	.677	.985**	1.000	.985**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.687	.985**	.702
	Sig. (2-tailed)	.	.016	.000	.000	.000	.035	.	.000	.000	.101	.	.000	.070	.140	.000	.	.000	.000	.043	.005	.005	.003	.421	.000	.000	.018	.000	.000	.000	.175	.000	.132	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
17	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
18	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
19	Correlation Coefficient	.826*	.636	.839*	.839*	.839*	.623	.826*	.839*	.839*	.906*	.826*	.839*	.892*	.516	.839*	.826*	.839*	.839*	1.000	.826*	.826*	.906*	.048	.839*	.839*	.633	.839*	.839*	.839*	.450	.839*	.798	.839*	.588
	Sig. (2-tailed)	.043	.175	.037	.037	.037	.186	.043	.037	.037	.013	.043	.037	.017	.294	.037	.043	.037	.037	.	.043	.043	.013	.929	.037	.037	.177	.037	.037	.037	.371	.037	.057	.037	.096
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

20	Correlation Coefficient	.939**	.803	.985**	.985**	.985**	.907*	.939**	.985**	.985**	.727	.939**	.985**	.896*	.800	.985**	.939**	.985**	.985**	.826*	1.000	1.000**	.864*	.318	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.627	.985**	.702
	Sig. (2-tailed)	.005	.054	.000	.000	.000	.013	.005	.000	.000	.101	.005	.000	.016	.056	.000	.005	.000	.000	.043	.	.	.027	.539	.000	.000	.018	.000	.000	.000	.175	.000	.183	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
21	Correlation Coefficient	.939**	.803	.985**	.985**	.985**	.907*	.939**	.985**	.985**	.727	.939**	.985**	.896*	.800	.985**	.939**	.985**	.985**	.826*	1.000**	1.000	.864*	.318	.985**	.985**	.890*	.985**	.985**	.985**	.636	.985**	.627	.985**	.702
	Sig. (2-tailed)	.005	.054	.000	.000	.000	.013	.005	.000	.000	.101	.005	.000	.016	.056	.000	.005	.000	.000	.043	.	.	.027	.539	.000	.000	.018	.000	.000	.000	.175	.000	.183	.000	.040
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
22	Correlation Coefficient	.955**	.773	.923**	.923**	.923**	.750	.955**	.923**	.923**	.864*	.955**	.923**	.806	.554	.923**	.955**	.923**	.923**	.906*	.864*	.864*	1.000	.364	.923**	.923**	.810	.923**	.923**	.923**	.572	.923**	.851*	.923**	.704
	Sig. (2-tailed)	.003	.072	.009	.009	.009	.086	.003	.009	.009	.027	.003	.009	.053	.254	.009	.003	.009	.009	.013	.027	.027	.	.479	.009	.009	.050	.009	.009	.009	.235	.009	.032	.009	.047
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
23	Correlation Coefficient	.409	.182	.369	.369	.369	.657	.409	.369	.369	.091	.409	.369	.134	.000	.369	.409	.369	.369	.048	.318	.318	.364	1.000	.369	.369	.715	.369	.369	.369	-.048	.369	.358	.369	.354
	Sig. (2-tailed)	.421	.730	.471	.471	.471	.157	.421	.471	.471	.864	.421	.471	.800	1.000	.471	.421	.471	.471	.929	.539	.539	.479	.	.471	.471	.110	.471	.471	.471	.929	.471	.486	.471	.524
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
24	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
25	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
26	Correlation Coefficient	.890*	.683	.904*	.904*	.904*	.984**	.890*	.904*	.904*	.572	.890*	.904*	.720	.581	.904*	.890*	.904*	.904*	.633	.890*	.890*	.810	.715	.904*	.904*	1.000	.904*	.904*	.904*	.433	.904*	.626	.904*	.677
	Sig. (2-tailed)	.018	.135	.013	.013	.013	.000	.018	.013	.013	.235	.018	.013	.107	.227	.013	.018	.013	.013	.177	.018	.018	.050	.110	.013	.013	.	.013	.013	.013	.391	.013	.183	.013	.060
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
27	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000	1.000**	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
28	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000	1.000**	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
29	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000	.645	1.000**	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

30	Correlation Coefficient	.636	.683	.645	.645	.645	.426	.636	.645	.645	.572	.636	.645	.564	.904*	.645	.636	.645	.645	.450	.636	.636	.572	-.048	.645	.645	.433	.645	.645	.645	1.000	.645	.423	.645	.601
	Sig. (2-tailed)	.175	.135	.166	.166	.166	.399	.175	.166	.166	.235	.175	.166	.244	.013	.166	.175	.166	.166	.371	.175	.175	.235	.929	.166	.166	.391	.166	.166	.166	.	.166	.404	.166	.221
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
31	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000	.667	1.000**	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
32	Correlation Coefficient	.687	.343	.667	.667	.667	.585	.687	.667	.667	.940**	.687	.667	.779	.364	.667	.687	.667	.667	.798	.627	.627	.851*	.358	.667	.667	.626	.667	.667	.667	.423	.667	1.000	.667	.644
	Sig. (2-tailed)	.132	.505	.148	.148	.148	.222	.132	.148	.148	.005	.132	.148	.068	.478	.148	.132	.148	.148	.057	.183	.183	.032	.486	.148	.148	.183	.148	.148	.148	.404	.148	.	.148	.178
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
33	Correlation Coefficient	.985**	.862*	1.000**	1.000**	1.000**	.889*	.985**	1.000**	1.000**	.739	.985**	1.000**	.849*	.750	1.000**	.985**	1.000**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	.904*	1.000**	1.000**	1.000**	.645	1.000**	.667	1.000	.695
	Sig. (2-tailed)	.000	.027	.	.	.	.018	.000	.	.	.094	.000	.	.033	.086	.	.000	.	.	.037	.000	.000	.009	.471	.	.	.013	.	.	.	.166	.	.148	.	.065
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

Appendix 17:

Correlation Coefficients Native English Participants

Spearman's rho		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Average
1	Correlation Coefficient	1.000	.265	.359	.359	.359	.265	.359	.000	.265	.319	.478	.359	.265	.359	.309	.441	.441	.088	-.441	.359	.359	.359	.216	.359	-.290	.288
	Sig. (2-tailed)	.	.612	.485	.485	.485	.612	.485	1.000	.612	.538	.338	.485	.612	.485	.552	.381	.381	.868	.381	.485	.485	.485	.681	.485	.577	.542
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
2	Correlation Coefficient	.265	1.000	.985**	.985**	.985**	1.000**	.985**	.727	1.000**	.776	.677	.985**	1.000**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.985**	.636	.985**	.687	.647
	Sig. (2-tailed)	.612	.	.000	.000	.000	.	.000	.101	.	.070	.140	.000	.	.000	.043	.005	.005	.003	.421	.000	.000	.000	.175	.000	.132	.081
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
3	Correlation Coefficient	.359	.985**	1.000	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
4	Correlation Coefficient	.359	.985**	1.000**	1.000	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
5	Correlation Coefficient	.359	.985**	1.000**	1.000**	1.000	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
6	Correlation Coefficient	.265	1.000**	.985**	.985**	.985**	1.000	.985**	.727	1.000**	.776	.677	.985**	1.000**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.985**	.636	.985**	.687	.647
	Sig. (2-tailed)	.612	.	.000	.000	.000	.	.000	.101	.	.070	.140	.000	.	.000	.043	.005	.005	.003	.421	.000	.000	.000	.175	.000	.132	.081
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
7	Correlation Coefficient	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
8	Correlation Coefficient	.000	.727	.739	.739	.739	.727	.739	1.000	.727	.896*	.554	.739	.727	.739	.906*	.727	.727	.864*	.091	.739	.739	.739	.572	.739	.940**	.665
	Sig. (2-tailed)	1.000	.101	.094	.094	.094	.101	.094	.	.101	.016	.254	.094	.101	.094	.013	.101	.101	.027	.864	.094	.094	.094	.235	.094	.005	.165
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

9	**Correlation Coefficient**	.265	1.000**	.985**	.985**	.985**	1.000**	.985**	.727	1.000	.776	.677	.985**	1.000**	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.985**	.636	.985**	.687	.647
	Sig. (2-tailed)	.612	.	.000	.000	.000	.	.000	.101	.	.070	.140	.000	.	.000	.043	.005	.005	.003	.421	.000	.000	.000	.175	.000	.132	.081
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
10	**Correlation Coefficient**	.319	.776	.849*	.849*	.849*	.776	.849*	.896*	.776	1.000	.728	.849*	.776	.849*	.892*	.896*	.896*	.806	.134	.849*	.849*	.849*	.564	.849*	.779	.676
	Sig. (2-tailed)	.538	.070	.033	.033	.033	.070	.033	.016	.070	.	.101	.033	.070	.033	.017	.016	.016	.053	.800	.033	.033	.033	.244	.033	.068	.103
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
11	**Correlation Coefficient**	.478	.677	.750	.750	.750	.677	.750	.554	.677	.728	1.000	.750	.677	.750	.516	.800	.800	.554	.000	.750	.750	.750	.904*	.750	.364	.667
	Sig. (2-tailed)	.338	.140	.086	.086	.086	.140	.086	.254	.140	.101	.	.086	.140	.086	.294	.056	.056	.254	1.000	.086	.086	.086	.013	.086	.478	.178
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
12	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
13	**Correlation Coefficient**	.265	1.000**	.985**	.985**	.985**	1.000**	.985**	.727	1.000**	.776	.677	.985**	1.000	.985**	.826*	.939**	.939**	.955**	.409	.985**	.985**	.985**	.636	.985**	.687	.647
	Sig. (2-tailed)	.612	.	.000	.000	.000	.	.000	.101	.	.070	.140	.000	.	.000	.043	.005	.005	.003	.421	.000	.000	.000	.175	.000	.132	.081
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
14	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
15	**Correlation Coefficient**	.309	.826*	.839*	.839*	.839*	.826*	.839*	.906*	.826*	.892*	.516	.839*	.826*	.839*	1.000	.826*	.826*	.906*	.048	.839*	.839*	.839*	.450	.839*	.798	.520
	Sig. (2-tailed)	.552	.043	.037	.037	.037	.043	.037	.013	.043	.017	.294	.037	.043	.037	.	.043	.043	.013	.929	.037	.037	.037	.371	.037	.057	.119
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
16	**Correlation Coefficient**	.441	.939**	.985**	.985**	.985**	.939**	.985**	.727	.939**	.896*	.800	.985**	.939**	.985**	.826*	1.000	1.000**	.864*	.318	.985**	.985**	.985**	.636	.985**	.627	.650
	Sig. (2-tailed)	.381	.005	.000	.000	.000	.005	.000	.101	.005	.016	.056	.000	.005	.000	.043	.	.	.027	.539	.000	.000	.000	.175	.000	.183	.067
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

17	**Correlation Coefficient**	.441	.939**	.985**	.985**	.985**	.939**	.985**	.727	.939**	.896*	.800	.985**	.939**	.985**	.826*	1.000**	1.000	.864*	.318	.985**	.985**	.985**	.636	.985**	.627	.650
	Sig. (2-tailed)	.381	.005	.000	.000	.000	.005	.000	.101	.005	.016	.056	.000	.005	.000	.043	.	.	.027	.539	.000	.000	.000	.175	.000	.183	.067
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
18	**Correlation Coefficient**	.088	.955**	.923**	.923**	.923**	.955**	.923**	.864*	.955**	.806	.554	.923**	.955**	.923**	.906*	.864*	.864*	1.000	.364	.923**	.923**	.923**	.572	.923**	.851*	.564
	Sig. (2-tailed)	.868	.003	.009	.009	.009	.003	.009	.027	.003	.053	.254	.009	.003	.009	.013	.027	.027	.	.479	.009	.009	.009	.235	.009	.032	.088
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
19	**Correlation Coefficient**	-.441	.409	.369	.369	.369	.409	.369	.091	.409	.134	.000	.369	.409	.369	.048	.318	.318	.364	1.000	.369	.369	.369	-.048	.369	.358	.299
	Sig. (2-tailed)	.381	.421	.471	.471	.471	.421	.471	.864	.421	.800	1.000	.471	.421	.471	.929	.539	.539	.479	.	.471	.471	.471	.929	.471	.486	.556
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
20	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000	1.000**	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
21	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000	1.000**	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
22	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000	.645	1.000**	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
23	**Correlation Coefficient**	.216	.636	.645	.645	.645	.636	.645	.572	.636	.564	.904*	.645	.636	.645	.450	.636	.636	.572	-.048	.645	.645	.645	1.000	.645	.423	.584
	Sig. (2-tailed)	.681	.175	.166	.166	.166	.175	.166	.235	.175	.244	.013	.166	.175	.166	.371	.175	.175	.235	.929	.166	.166	.166	.	.166	.404	.243
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
24	**Correlation Coefficient**	.359	.985**	1.000**	1.000**	1.000**	.985**	1.000**	.739	.985**	.849*	.750	1.000**	.985**	1.000**	.839*	.985**	.985**	.923**	.369	1.000**	1.000**	1.000**	.645	1.000	.667	.647
	Sig. (2-tailed)	.485	.000	.	.	.	.000	.	.094	.000	.033	.086	.	.000	.	.037	.000	.000	.009	.471	.	.	.	.166	.	.148	.102
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000
25	**Correlation Coefficient**	-.290	.687	.667	.667	.667	.687	.667	.940**	.687	.779	.364	.667	.687	.667	.798	.627	.627	.851*	.358	.667	.667	.667	.423	.667	1.000	.613
	Sig. (2-tailed)	.577	.132	.148	.148	.148	.132	.148	.005	.132	.068	.478	.148	.132	.148	.057	.183	.183	.032	.486	.148	.148	.148	.404	.148	.	.187
	N	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6.000

Appendix 18:

Participants comments on the CAP-II Study

Participant 1 (Pilot)

- Videos 1 and 2 are both very clear.
- Video 4 not as clear as the other videos this made it more difficult to decide on a CAP-II score.
- Video 4 is the most difficult video to judge.
- The order of the CAP-II rating scale may not be clear cut.
- There may be need for more categories in the CAP-II

Participant 2 (Pilot)

- Video 1 was viewed partly again and re-rated.
- Video 2 and 5 was viewed again and re-rated.
- Overall context of the experiment was not clear.
- It is unclear what CAP-II category that you need to tick.
- Need to say to participants that they need to tick the highest level.
- You may want to rate some or all of the videos the same.
- Queried the order of CAP-II 8 and 9 – what if the participant can score a 9 when the child cannot do 8?
- The additional comment for 9 is useful.
- The additional comment for 8 is too long winded.

<u>Extra changes to make based on participant 2</u>

- Explain to subjects that they need to tick the hardest category.
- State that they may want to repeat.
- State that it is ok to rate the videos the same.

Participant 3 (Pilot)

- It is good to start each video yourself and have time to do it. (You get time to think which category to select).
- Participant was indecisive a couple of times.
- Video 4 – Background noise is a problem / language delay –telephone use is variable.
- Video 4 – In quiet child is at her best when using telephone (CAP-II = 7) / In noise (CAP-II = 4).
- Video 5 – telephone uses the speaker setting.
- Video 1 – conversation but with filling in sometimes.

Participant 4

- Video 3 the interviewer asks if child can use the phone with an unknown speaker in noise – Participant states that this is not the question on the scale.
- Video 6 for the same CAP-II category the interviewer asks the same question only in quiet, there is no mention of noise – different situation.
- Found a problem with video 3 – didn't ask about CAP-II 9. Asked about phone in noise.

Participant 5

- Found the experiment interesting.

Participant 6

- Needed one or two things on the CAP-II rating scale clarified.
- Video 1 was confusing – the interviewer missed out some of the scale, it was unclear. Questioned whether or not the child can do CAP-II 6.
- Video 1 the interviewer didn't ask about the easier task just the harder task.
- Child cannot do CAP-II 6 but he can do CAP-II 7- so rated CAP-II 5.
- Participant thinks there is something wrong with the scale.

Participant 7

- Participant thinks that video 1 and video 3 are clearer than the other videos.
- Participant found video 4 more difficult to judge – on side of counter participant would rate at a 6 but the child seemed able to partly fulfil CAP-II 7, so rated 7.
- Participant rated video 5 a CAP-II 8 although the child may not be able to do this 100% of the time.
- Participant states that video 5 would be a 7 with use of a speaker attached to the phone.

Participant 8

- Video 2 - the child uses the phone with unknown speakers but sometimes needs clarification.
- Video 1- participant states that category 6 maybe harder for the child than category 7.
- Video 5 - it is unknown whether or not CAP-II 9 is possible for the child as there is no knowledge about this category due to the age of the child.
- Participant feels it would be better to state in CAP-II 9 about using the phone with an unknown speaker in noise rather than in an unpredictable context.
- Current CAP-II 9 is easier if it was in quiet as opposed to in noise.

Participant 9

- Participant feels that CAP-II 8 is confusing as sometimes in the video it states with lip-reading and in other videos it doesn't. (It is re-phased wrong).
- Participant queried this for video 5 – without lip-reading is harder than with lip-reading.
- Participant feels that the interviewee bases a lot of her responses from what the child says to her in video 3. Participant feels it would be good to have the interviewee's own observations of the child, as the child maybe biased especially if he wants to be 'cool'.

- Video 1- the child's language could have been elaborated. It was said having a conversation without lip-reading would require someone 'filling in'. Participant queried whether this was good enough, he was unsure whether to rate CAP-II 6 or 7.

Participant 10

- Participant finds the experiment interesting.

- Video 1 - participant was unsure how to rate as the child can do a level higher than the previous level which they cannot do. I.e. The child can do CAP-II 7 but not CAP-II 6.

- Telephone categories are not relevant because the CAP-II is mainly used to assess young children.

- It does not state whether CAP-II 7 and 9 are in background noise. (This was only stated in one video, video 3.

- CAP-II 9 children usually do not come into that circumstance.

- Video 4 - the child couldn't understand the telephone in a noisy background and the criteria didn't incorporate this.

- Video 6 - child picked up the phone but they didn't specify whether they could totally understand because of brief conversation.

- Video 5 - Even though the child is using the telephone via a speaker system it is not totally understood whether she can use a normal telephone. And the child doesn't us it much.

- Video 1 – CAP-II 7 was given although it seems that the child could not fully understand a conversation without lip-reading.

- Video 3 – CAP-II 8 was chosen because at the end the interviewee said that the child may only use the phone with a known listener in noisy background it is difficult.

- Video 2 – sometimes it is unknown whether the child can understand an unknown speaker.

Participant 11

- Queried terminology – decrements in video 4.
- Writing notes throughout video 5, 2, 3.
- Would like to see the child.
- Specify CAP-II 9 more either in noise or quiet or include both.

Participant 12

- No comments

Participant 13

- Video 1 - the child can use the telephone with a known listener but can't understand a full conversation so CAP-II 5 was given. Participant struggled to know which category to give.
  - Video 4 - the child can do tasks in noise in a controlled testing situation but parents say that she cannot manage as well if she was at the zoo for example. Conflicting evidence, parents and testing in background noise 2 different responses.
  - Video 2 – The interviewer does not stipulate whether conversation is with or without lip-reading (CAP-II 6).
  - Sometimes when questions are asked in a different order it can make it more difficult. I.e. Video 5 – jumps from CAP-II 6 to 8 back to CAP-II 7. Participant doesn't know where to tick as the child can do 8 but not 7 i.e. a speaker phone is used and the child can do 6.
  - Video 2 - doesn't jump about when questioning.

Participant 14

- The participant thinks that some of the CAP is too general it needs to be more specific.
- Some child are doing better at CAP-II 9 but not as well in CAP-II 8 so perhaps they should be reordered.
- The videos are very clear.

- In some cases the interviewees didn’t know how the child was doing and they were guessing.

Participant 15

- The whole experiment had to be repeated!
- Participant watched the videos in the wrong order and had those all numbered wrong.
- The participant thinks that the judgement of each child is based on the interviewees’ judgement rather than her own.
- Video 1 – CAP-II 6 seemed slightly unclear.
- Video 5 – not much knowledge of telephone because of age.

Participant 16

- No comments

Participant 17

- Videos are very clear.
- Suggests maybe including a category harder that category 9.

Participant 18

- Video 1 – Participant took a while to rate this, she seemed unsure.
- Video 4 – queried terminology decrements.
- Asked environmental sounds to be clarified i.e. the difference between CAP-II 1 and 3.
- Scale can be inconsistent – 2 people can score the same i.e. 7 but one child could be performing a lot better than another person who has the same score.
- The scale doesn’t take into account the child’s personality or what is expected of a child at a particular age i.e. confident 6yr old will answer the phone but a shy 6 yr old may not.
- How long is the child implanted?

- Participant thinks that CAP-II 8 and 9 are very useful and that they are very different.
- CAP-II 6 and 9 the personality of the child needs to be considered.
- Video 6 – the child is able to have a long conversation on the phone with an unknown listener – not covered by the scale.
- Video 1 found it difficult to choose either CAP-II 5 or 7 as the child could not do CAP-II 6.
- Video 4 – Would give a CAP-II 7 if the child can use the phone in quiet but would only give a 4 if it was in a noisy environment.
- Video 5 – child has not been tested for CAP-II 9.

Participant 19

- Video 5 – It is not clear what levels of conversation in a classroom the child is able to follow. The child can join in but specifics are difficult to discern from questioning.
- Video 4 – Category 7 depends on the child's emotional state so perhaps some doubt about competence overall in this category.
- Video 6 – Not sure if background noise would pose a problem here.
- Video 2 – question was answered by interviewer prompting, subject to bias – cannot be fully confident at level 9.
- Video 3 – was quite clear.
- Video 4 – participant seemed slightly unsure.
- Can be difficult to rate as questioning is unpredictable.
- Too much positive prompting from the interviewer this may make the answers of the interviewing a bit biased.
- Questions were not asked in a controlled way.

- The easiest videos to rate where those were the interviewee was very positive and they knew the child well. Videos 2, 3 and 6.
- Interesting experiment.
- Queried CAP-II 7 – how many known listeners 1 or more? Hard to define.
- CAP-II 8 - could be doubt about the child's ability to listen in background noise as some say yes but need repeating.

Participant 20

- Found it interesting.
- Video 5 – the child maybe capable of category 9 but too young, not been explored.
- Video 1 – The child can do 7 but not 6 so they say they cannot tick 7.
- The child has additional language difficulties that may be a factor when using the phone.

Participant 21

- Video 4 – watching it a second time helped the participant decide.
- Participant asked questions – seemed slightly unsure about what category to choose for video 3 (This was the first video watched).
- Needed the experiment re-explained as there was slight confusion over the scale.
- Video 1 - was re-rated as a result of explanation.
- After video 1 and the instructions were re-explained participant was able to continue reliably.
- Interesting experiment.

Participant 22

- Some background noise.
- Interviewer asked the questions well.

- The rating scale is easy to use.
- Few interruptions.

Participant 23

- No comments.

Participant 24

- No comments.

Participant 25

- Quite clear.

Participant 26

- Asked about terminology – decrements in video 4.
- Re-rated video 6 – from CAP-II 9 to 8.
- Video 2 – was the first video rated was unsure what to rate as the child could do everything, then rated CAP-II 9.

Participant 27

- Video 5 – child is too young for CAP-II 9.
- Straightforward.

Participant 28

- Background noise
- Video 4 – using the telephone is difficult as child has language problems.
- Video 5 – the child is young so all categories have not been assessed.

Participant 29

- Video 5 – child is only getting partial info but the video seems to suggest the child is able to have conversation in a classroom.
- Videos and sound is very clear.
    - Questioning is clear – asking specific questions and answering specific questions.

- Professionals descriptions may be incorrect the questions are general.
- The professionals could be answering in response to one case but it may not apply to all cases in general.

Participant 30

- Video 3 - difficult to rate.
- Found that the child may fulfil one category but not the one above.
- Found that some categories are not fully fulfilled – this made it more difficult to decide what to rate. The participant decided to rate the highest level.
- Video 3 – it was not mentioned whether the child was able to have a conversation with a known listener but they did say that he is able to follow group conversations – they didn't clarify whether in background noise though.
- Video 1 –Professionals stated that patient is able to have a conversation with a familiar listener so the participant rated it a CAP-II 7. However it was not stated clearly that he is able to understand a full conversation without lip-reading.

Participant 31

- Video 1 - the participant was unsure whether the child could do CAP-II 6 fully but she rated 7 as they could do 7.
- Participant feels that phone questions are not always suitable for young children.
- Video 5 – the participant ticked 8 but she is unsure if the child can do 7.
- The participant felt that the questioning was too influential – the questions were not asked in an unbiased way.

Participant 32

- Interesting
- Video 4 - was stopped before the end (If CAP-II low this is the reason).

- Participant thinks that the scale is good and that you could tick more than one category.

Participant 33

- Video 4 – participant ticked CAP-II 7 but states that it only applies when the child is calm and in quiet.
- Video 5 – unknown due to the child being so young.
- Clarification was needed between common phrases and conversation. The differences between CAP-II 5 and 6 were defined.
- Participant feels that the interviewer is quite leading with the questioning – could be putting the answer into the interviewees head.
- Telephone questions should maybe be put on a separate scale.
- Have the possibility of asking CAP-II 7 after 8.
- Scale follows through well but not in terms of age appropriateness.
- Have the possibility of N/A next to CAP-II 7 and 9.
- People selecting CAP-II 8 my not indicate that the child can definitely do 7.
- The last 3 categories do not follow as well.
- Questioned whether CAP-II 8 is with or without lip-reading.
- Videos are covered well.
- CAP-II 7, 8 and 9 are too simplistic.
- Other factors to do with the phone - not applied well enough to every situation.
- It is not stated if using their Cochlear implant with the phone/new implant or old one.

General comments

Category 9 is discussed before 8 in video 6. The majority of subjects ticked CAP-II 9. However some people ticked CAP-II 8. (A possible reason for this could be the order that the listening skills are discussed).