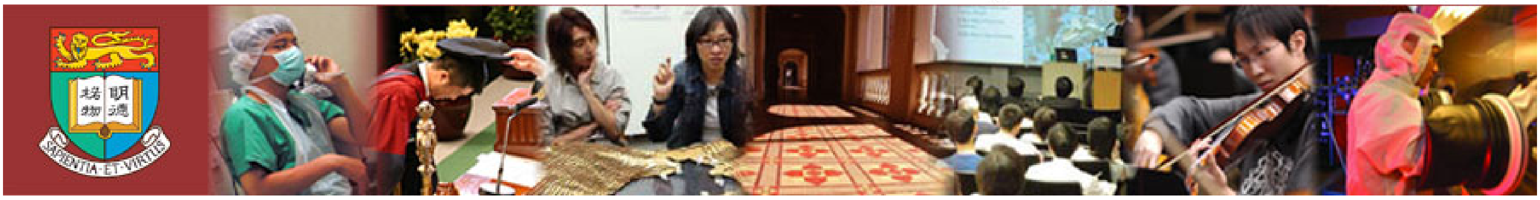


Otoscopic visualization of cerumen: Inter-rater agreement

Chan Hin Yau, Denise

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, May 6, 2005

Abstract

The study investigated the inter-rater agreement of Sullivan scale in classifying the amount of cerumen. Undergraduates from the Division of Speech and Hearing Sciences were recruited as raters, and received training on otoscopy. In the rating program, raters had to rate the subject ears with the Sullivan scale using hand-held otoscope and video-otoscope. It was found that the inter-rater agreement was fair for hand-held otoscopy and moderate for video-otoscopy. Moreover, the level of inter-rater agreement varied with the amount of cerumen, with the end points of the scale possessed a higher percentage of agreement than the mid points. Results showed that Sullivan scale would be a suitable means for quantifying the amount of cerumen. Video-otoscopy could be regarded as the preferred primary procedure rather than hand-held otoscopy in rating the amount of cerumen. Finally, Sullivan scale may have more utility if the definitions of the mid points are refined.

Otoscopic visualization of cerumen: Inter-rater agreement

Cerumen (“ear wax”) refers to a mixture of the secretary products of the sebaceous and ceruminous glands in the auditory canal, as well as the desquamated sheets of corneocytes (keratin squamae) originating from the canal wall. Cerumen can be classified into dry and wet types. Dry cerumen is yellow or grey, flaky and brittle, while wet cerumen is brown, wet and sticky. Wet cerumen is either soft or hard. Soft cerumen is characterized by its moist and sticky texture and small sheets of keratin squamae. Hard cerumen, on the other hand, is dry and desiccated, with much large, dense and prominent keratin squamae (Guest, Greener, Robinson & Smith, 2004; Hawke, 2002). An individual’s cerumen type is hereditary, with the allele encoding the wet form being dominant over the dry form (Ibtaimov, 1991).

A moderate amount of cerumen is crucial in maintaining healthy ears. Cerumen basically serves three purposes: it lubricates the canal; it cleans the canal; and it protects the canal against bacterial and fungal inflections. In normal individuals, excessive cerumen is removed by a mechanism called migration—the superficial keratin squamae move laterally from the tympanic membrane to the outer ear. However, in some individuals this mechanism does not function properly, most commonly caused by a failure of corneocyte separation (Hawke, 2002), resulting in cerumen accumulation. Excessive cerumen accumulation in the auditory canal can adversely affect auditory functions, and the degree of interference changes with its amount, texture, level of impaction and position within the canal. Usually, the adverse effect is directly proportional to the quantity of cerumen. Hard cerumen is the commonest type to be linked with chronic and recurrent cerumen impaction. Finally, the more impacted the cerumen and the shorter the distance between the impacted cerumen and the tympanic membrane, the higher the resistance to removal.

Many investigations reported an association between cerumen accumulation and hearing dysfunctions. It was found that 30% of elderly patients with hearing loss had cerumen impaction as a fundamental cause (Lewis-Culinan & Janken, 1990). Cerumen accumulation leads to conductive hearing loss by physically obstructing sound transmission, which may decrease hearing acuity by 40-45 dB (Meador, 1995). Moreover, high frequency signals were particularly prone to the deleterious effects of occlusion. Cerumen impaction is not restricted to the elderly, and had been found in people of all age groups (Guest et al., 2004). A recent study has shown that children with a history of cerumen accumulation were more likely to have hearing loss, and unexpectedly, a significant association was found between impacted cerumen and mixed hearing loss (conductive and sensorineural hearing loss), apart from that between impacted cerumen and conductive hearing loss (Olusanya, 2003). Hearing impairment in turn may influence an individual’s emotional, psychological and social functions, as well as their quality of life (Mulrow et al., 1990). Besides, it might have a deleterious effect on the cognitive and academic development of children. Apart from hearing impairment and its secondary influence, impacted cerumen might lead to a variety of symptoms including perforated eardrum, itching, pain, tinnitus, dizziness and increased risk of infection (Burkhart, Kruge, Burkhart & Black, 2001). Moreover, cerumen accumulation may limit audiological access to the ear for procedures such as impression taking, hearing aid fitting, probe-tube microphone measurement, tympanometry and other audiological measures. To limit the damaging effects of excessive cerumen, cerumen management is a necessity.

Impacted cerumen is physically curable. It can be removed by irrigation, suction, cerumen removal implements or a combination of the above methods. Surgery may be appropriate as well but is necessary only in rare cases. Research has always found a positive outcome for hearing ability following cerumen removal, with the majority of patients reporting an increase in hearing acuity (Lewis-Culinan & Janken, 1990; Moore, Voytas, Kowalski & Maddens, 2002). An average of 5 dB improvement was reported following occlusive cerumen removal in one study (Sharp, Wilson, Ross & Barr-Hamilton, 1990).

The question that this research study is concerned with is how to determine if a certain amount of cerumen is optimal or excessive, and in turn the necessity of its removal. This is a very practical question in the speech therapy clinic. Speech therapists often refer clients with obvious visual signs of ear disorder to medical practitioners or audiologists for detailed assessment and management. Very often, the observed problem is impacted cerumen. An answer to this question could provide helpful guidance to the referring clinician.

Speech therapists often use otoscopy to investigate the external ear and the tympanic membrane (ASHA, 2001). It is the primary method of evaluating the amount of cerumen present and an important skill for diagnosing external and middle ear conditions. Otoscopy is an easy-to-carry-out technique that requires few instruments, namely an otoscope and an attached speculum. The otoscope projects light into the auditory canal, which allows visualization of the canal as well as the tympanic membrane. The speculum is used as an ear-fitting device. There are various types of otoscopes, including the hand-held otoscope, the video-otoscope and the pneumatic otoscope. The hand-held otoscope is the simplest instrument. The video-otoscope, with a video camera incorporated into the otoscope optical system, digitalizes the images of the auditory canal and the tympanic membrane, and allows the otoscopic images to be reviewed, stored, archived, and transmitted for medical specialist opinion. Documentation of treatment outcomes is easily attained by comparing the pre- and post-treatment images. Moreover, the video-otoscope provides a unique means to educate and counsel patients when they are allowed to view their own ears (Sullivan, 1997). The pneumatic otoscope, on the other hand, is specifically designed to assess the mobility of the tympanic membrane and the middle ear space as both positive and negative pressure can be delivered through the otoscope using a pneumatic bulb. Pneumatic otoscopy is mainly performed by medical doctors. Each kind of otoscope has been developed for different functions, and selection depends on the purposes of the assessment.

An accurate otoscopic diagnosis is dependent on the practitioner’s experience and training (Wormald, Browning & Robinson, 1995). Clinical experience is accumulated over years, while training needs to include how to justify the amount of cerumen accumulation and when to recommend cerumen removal. Thus, diagnostic criteria should be made explicit. In 1995, Sullivan had developed a scale which was used to measure the level of cerumen obstruction of the auditory canal (Appendix A). The Sullivan scale, designed for use with a video-otoscope, is a 4-point scale for classification of cerumen accumulation. It ranges from 0 (cerumen entirely absent) to +3 (an occlusive, major amount of cerumen present).

The Sullivan scale provides a systematic way to quantify and qualify the amount of cerumen. Cerumen removal is usually suggested for ears with the ratings +2 and +3. Nevertheless, the Sullivan scale is a subjective rating method. To the author’s knowledge, there has been no study investigating its inter-rater agreement, reliability or accuracy in diagnosis. This greatly limits the utility of the scale.

The subjectivity of the Sullivan scale leads to the first purpose of this study, to verify its inter-rater agreement in evaluating the amount of cerumen accumulation. Hand-held otoscope and video-otoscope procedures are chosen in the investigation. The hand-held otoscope is the commonest instrument available in most speech clinics as it is more affordable. In addition, it is easier to use and portable. However, the hand-held otoscope often shows only a portion of the tympanic membrane, because of the limited size of the speculum and the long distance between the objective lens and the ear speculum. The video-otoscope, on the other hand, provides a relatively high quality image which is larger, clearer and well-focused. Another advantage is that the entire tympanic membrane can be visualized. A point to note is the Sullivan scale was primarily designed for use with video-otoscope, and its use with a hand-held otoscope in this study is an extension of the original application of the scale. The second purpose of this study is to investigate whether better inter-rater agreement is obtained from the video-otoscope procedure. The video-otoscope can be used to substitute or supplement the preliminary standard otoscopic examination in the initial stage of most audiological consultations, because of its greater clarity and detail (Sullivan, 1993). Many speech therapists who work in proximity to audiologists also have access to video-otoscopy facilities. This study will consider if the video-otoscope is equivalent, or superior to, the hand-held otoscope in terms of inter-rater agreement when used by speech therapy students. In addition, how does the inter-rater agreement vary will be investigated.

Therefore, the main research questions involved in this study were:

1. What is the inter-rater agreement of the Sullivan scale when using a hand-held otoscope?
2. What is the inter-rater agreement of the Sullivan scale when using a video-otoscope?
3. Does a video-otoscope give better inter-rater agreement than a hand-held otoscope?
4. Does the level of inter-rater agreement change as a function of the amount of cerumen accumulated?

Method

Raters

Seventeen raters (16 female and 1 male, age range between 18 to 25, with a mean age of 22 years) participated in the rating program. All raters are undergraduates from the Division of Speech and Hearing Sciences, The University of Hong Kong. All raters reported to have normal hearing and normal visual acuity (either without spectacles/contact lens or when wearing spectacles/contact lens).

All raters had attended a 3-hour training workshop three days in advance of the rating session. The workshop was held by the research supervisor (Dr. Bradley McPherson) of this study (the workshop outline can be referred to in Appendix B). The first part was a lecture in which raters were taught about the anatomy of the ear, common disorders found at the outer ear, otoscopy, as well as the Sullivan scale. Photographs depicting ears with different levels of obstruction along the Sullivan scale were shown to the raters. The photographs were adapted from those on a website written by R. F. Sullivan, the scale's developer. The second part was a practical session. Raters were required to rate a series of model ears with simulated cerumen deposits by video-otoscope. In addition, the raters were required to rate the ears' cerumen status among each other by using hand-held otoscopes. Their ratings were compared to the anchor values subsequently provided by the research supervisor. Seminar handouts had been given to the raters. A question and answer session concluded the training workshop.

Ear Sample

Forty-four voluntary subjects (10 male and 34 female, age range from 61 to 94, with a mean age of 75 years) had been enrolled in this study and all of them were members of the Caritas Central District Elderly Center, Hong Kong.

All the subjects were required to attend an auditory status screening session conducted by the researcher of this study prior to the rating program. The subjects' ears were screened by the hand-held otoscope. This was to ensure the subject ears recruited were pathology free and included ratings judged by the researcher to be distributed fairly evenly between 0 and +3 of the Sullivan scale. These procedures helped to eliminate other confounding visual factors and to prevent the final inter-rater agreement calculations from bias towards one end of the scale. Moreover, to reduce possible rater bias effects, only subjects' right ears were used.

Thirty-two subjects (8 male and 24 female, age range from 61 to 94, with a mean age of 77 years) passed the screening and consented to participate in the rating program (the consent form can be located in Appendix C). The subjects were distributed evenly into two groups, the hand-held otoscopy group (n = 16) and the video-otoscopy group (n = 16).

Materials

Four hand-held otoscopes (Welch Allyn Model 25020A) and a video-otoscope (The Professional Video Otoscope, Intermedia Communications, California) were used in the rating program. Speculae of different sizes were prepared to fit the patient ears. A recording sheet had been provided to the raters to mark their results (Appendix D).

Procedures

All measurements took place at a meeting room in the Caritas Central District Elderly Center. The meeting room was well illuminated.

The rating program was divided into sixteen sessions, and each session took about 25 minutes. Within each session, one hand-held otoscopic measurement and one video-otoscopic measurement were made. Each ear was imaged in the same session by all raters. This was to guarantee a consistent ear status was viewed by all raters within a particular time frame. The hand-held otoscopic rating was carried out individually. To reduce any order-effect, the seventeen raters took turns to rate the ears in a random starting order with subsequent rotation. In contrast, the video-otoscopic measurement was carried out jointly. The researcher was responsible for capturing the image while raters took the ratings by looking at the video-otoscopic screen simultaneously. The schedule and the work-flow of the measurements can be referred to in Appendix E. All raters were asked to rate the amount of cerumen independently. Moreover, they were asked not to communicate the otoscopic results with each other during the measurement. The following were guidelines in the measurements: (1) Ensure the room illumination is good; (2) Turn on the illuminating switch of the otoscope; (3) Fit the otoscope with a clean and appropriate sized speculum; (4) Pull slightly on Darwin's tubercle backward and upward; (5) Insert tip of the speculum into the auditory canal. Be careful not to touch the canal wall; (6) If the tympanic membrane cannot be viewed, change the viewing angle or change the size of the speculum if necessary; (7) Record the rating result

Evaluation questionnaire

After the rating program, all the raters were asked to complete a questionnaire (Appendix F). The questionnaire concerned the raters' self-rated confidence in operating the otoscopes, in identifying the images of the tympanic membrane and cerumen, and in rating by the Sullivan Scale. Responses were graded on a Likert scale from 1 to 5, with 1 indicating "not confident at all" and 5 indicating "very confident" (Table 1).

Table 1

Student self-confidence level rating scale

Confidence rating	1	2	3	4	5
Confidence level	Not confident at all	Slightly confident	Somewhat confident	Confident	Very Confident

Results

Table 2 shows the distribution of the ratings in the hand-held otoscopy group and video-otoscopy group, respectively.

To determine whether individual raters gave similar ratings for the ears investigated, the first-order agreement coefficient (AC_1 statistic) was utilized (Gwet, 2001a). The AC_1 statistic was chosen because it is applicable to any reliability study where ratings are made on a

Table 2

Distribution of Otoscopy Ratings using the Sullivan Scale

Panel A: *The distribution of ratings in the hand-held otoscopy group*

Subject		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sullivan scale	0	8	5	0	10	1	3	17	1	0	0	16	4	0	0	13	2
	+1	7	8	4	4	5	11	0	6	9	0	1	10	2	12	4	12
	+2	2	4	9	3	7	3	0	8	7	3	0	3	15	5	0	3
	+3	0	0	4	0	4	0	0	2	1	14	0	0	0	0	0	0
Total		17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17

Panel B: *The distribution of ratings in the video-otoscopy group*

Subject		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sullivan scale	0	1	3	1	0	14	10	2	0	0	0	16	0	17	0	0	0
	+1	14	3	5	0	3	7	15	3	7	0	1	0	0	0	1	11
	+2	2	11	6	3	0	0	0	13	10	12	0	2	0	9	14	6
	+3	0	0	5	14	0	0	0	1	0	5	0	15	0	8	2	0
Total		17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17

categorical scale. It is equally valid when there are multiple response categories or when there are multiple raters. In addition, the AC_1 statistic possesses a well-defined theoretical framework that describes clearly how the coefficient is estimated, which facilitates the generalization of the results. A well-thought theoretical framework is necessary in assessing the inter-rater agreement because many other statistical tests, e.g. the Kappa statistic, focus extensively on the calculation procedures at the expense of a detailed supporting theoretical construct, which can lead to difficulties in the interpretation of results (Gwet, 2001a).

The AC_1 statistic between multiple raters (r), is defined as the conditional probability that two randomly selected raters agree given that there is no agreement by chance. The AC_1 statistic ranges from 0 to 1. A "0" indicates that all the ratings are random while a "1" indicates the smallest possible chance-agreement probability. In another words, the inter-rater agreement is considered as "good" when AC_1 statistic is approaching "1" and as "poor" when AC_1 statistic is approaching "0". To facilitate the interpretation of the AC_1 statistic, criteria are made for describing the level of agreement, based on those typically used with statistics of agreement (Landis & Koch, 1977) (Table 3).

Table 3

Criteria of the AC_1 Statistic

AC_1 Statistic	0.00-0.09	0.10-0.20	0.21-0.40	0.41-0.60	0.61-0.80	0.81-1.00
Descriptors	Poor	Slight	Fair	Moderate	Substantial	Perfect

The AC_1 statistic between r raters is given by $(P\alpha - Pe.\gamma) / (1- Pe.\gamma)$ where $P\alpha$ and $Pe.\gamma$ are the agreement probability and chance-agreement probability, respectively. Agreement probability is defined as the probability that two raters randomly selected from the pool of r raters agree after rating a subject randomly selected form the population of subjects. Chance-agreement probability is defined as the probability that two raters selected randomly from the pool of r raters agree by chance. The results were summarized in Table 4.

Table 4

Inter-rater agreement (AC_1 statistics) for the Sullivan Scale

	$P\alpha$	$Pe.\gamma$	$AC_1 = \frac{(P\alpha - Pe.\gamma)}{(1 - Pe.\gamma)}$	Significance level
Hand-held otoscopy	0.52	0.24	0.37	$p < .01$
Video-otoscopy	0.62	0.25	0.50	$p < .01$

The AC_1 statistic was 0.37 for hand-held otoscopy. The AC_1 statistic was 0.50 for video-otoscopy. The level of inter-rater agreement was considered as fair for hand-held otoscopy and as moderate for video-otoscopy groups, at $p < .01$ significance level (Gwet, 2001b). Video-otoscopy had a higher level of inter-rater agreement than hand-held otoscopy.

Figure 1 and 2 visually represent the distribution of otoscopy ratings, and is based on information presented in Table 2. It was observed that the level of inter-rater agreement varied extensively among rated ears. In order to know the pattern of the distribution of the ratings, the percentage of agreement for each scale category was therefore calculated.

The percentage (%) of agreement was defined as the proportion of raters with a majority selected rating category (the rating category that the majority of raters agree with for a

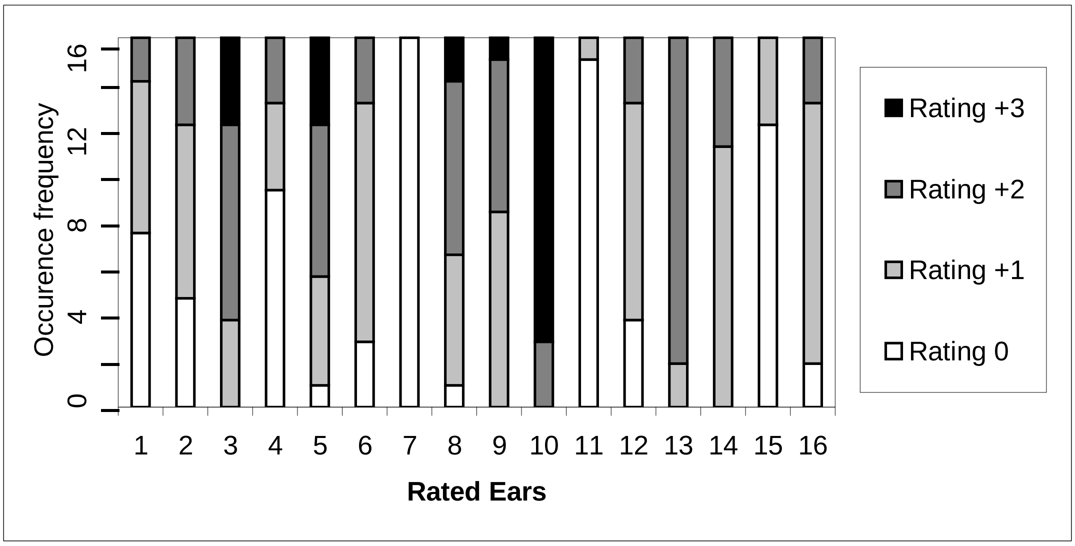


Figure 1. The distribution of ratings in the hand-held otoscopy group

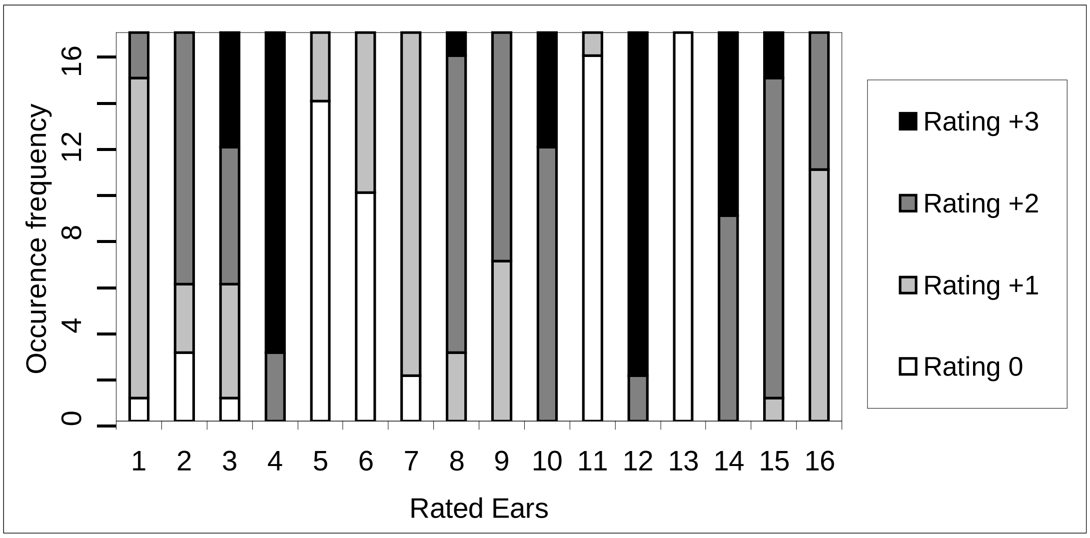


Figure 2. The distribution of ratings in the video-otoscopy group

particular ear) to the total number of raters. This analysis sought to determine if the % of agreement differed among different Sullivan scale categories, i.e., 0, +1, +2, +3 (Table 5).

The % of agreement ranged from 41% to 100% in the hand-held otoscopy group and from 35% to 100% in the video-otoscopy group. In both groups, the highest % of agreement was found in the category 0 while the lowest % of agreement was found in the category +2.

Table 5

The Percentage of Agreement for each Scale Category. The bracket "( )" refers to the overall frequency that the rating has been selected by the majority of raters

	% of agreement	Sullivan Scale			
		0	+1	+2	+3
Hand-held Otoscopy	Maximum	100%	71%	88%	82%
	Minimum	47%	47%	41%	82%
	Mean	75% (5)	61% (6)	57% (4)	82% (1)
	Overall	62% [End points*=76%; Mid points¶=59%]			
Video-otoscopy	Maximum	100%	88%	82%	88%
	Minimum	59%	65%	35%	82%
	Mean	81% (4)	78% (3)	63% (7)	85% (2)
	Overall	73% [End points*=82%; Mid points¶=68%]			

Note. *End points = 0 and +3; ¶Mid points = +1 and +2

When the mean % of agreement was compared, the end points of the scale, i.e. 0 and +3, possessed a higher value than the mid points of the scale, i.e. +1 and +2, for both otoscopy groups. The differences between the end and the mid points were 17% (76% - 59%) in hand-held otoscopy and 14% (82% - 68%) in video-otoscopy. The scale categories arranged in descending order in terms of the grand mean % of agreement were +3, 0, +1, +2.

The overall % of agreement was higher in the video-otoscopy group than in the hand-held otoscopy group. In addition, the mean % of agreement for each scale category was consistently higher for video-otoscopy than hand-held otoscopy ratings. The % difference ranged from 3% to 17% (mean = 8%), while category +1 possessed the largest difference (78% - 61% = 17%) and category +3 the smallest difference (85% - 82% = 3%).

The level of inter-rater agreement was also analyzed within subgroups based on the raters' self-rated confidence towards otoscopy. Raters were divided into 2 groups, the confident group and the less confident group. The confident group was made up of raters whose self-ratings' ranging from 4 to 5 ('confident" to "very confident") while the less confident group was made up of raters whose self- ratings' ranging from 1 to 3 ("not confident at all" to "somewhat confident"). The raters' self-rated confidence towards different parameters about otoscopy could be referred to in Appendix G.

A higher proportion of raters were classified into the confident group than the less confident group. The majority of the responses (47%; 72/153) were found in rating level 4, "confident". Only 0.01% (1/153) of all responses was at rating level 1, "not confident at all".

For the hand-held otoscopy group, self-confidence was the highest for operating the instrument and in identifying cerumen while the lowest was for rating actual client ears. For the video-otoscopy group, self-confidence was highest in identifying cerumen while the lowest was in operating the instrument. The distribution of confident raters and less confident raters in the two otoscopy groups were quite different in several aspects. For example, there were three-fold more confident raters in operating the hand-held otoscope than for operating the video-otoscope; there were two-fold more confident raters in rating actual client ears by using the video-otoscope than for rating use of the hand-held otoscope.

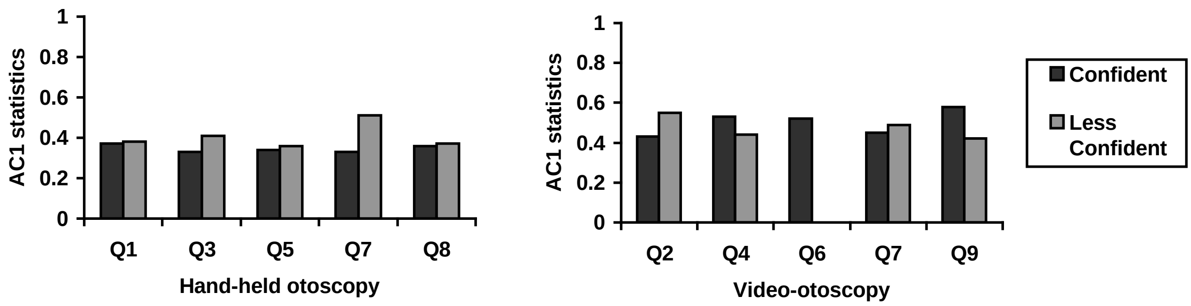


Figure3. Comparison of AC_1 statistic upon raters' self-confidence

The AC_1 statistics for the confident group and the less confident group were compared (Figures 3). It was observed that a higher self-rated confidence level did not yield a higher AC_1 statistic for hand-held otoscopy group in any self-rated parameters. In fact, for some of the parameters, such as the identification of tympanic membrane (Q3) and rating by Sullivan scale (photo/picture) (Q7), the less confident group attained a considerably higher AC_1 statistic than the confident group. Concerning the video-otoscopy group, two parameters, the operation of video-otoscope (Q1) and rating by Sullivan scale (photo/picture) (Q7), also possessed inverse results, while for another two parameters, the identification of tympanic membrane (Q4) and rating by Sullivan Scale (actual client ears) (Q9), the confident group had a higher AC1 statistic than the less confident group.

Discussion

The purpose of this study was to determine if independent raters would arrive at similar decisions using the Sullivan rating scale for the amount of cerumen. It was found that the level of inter-rater agreement was fair and moderate for hand-held otoscopy and video-otoscopy, respectively. Given that there were no other published means/scales to quantify the amount of cerumen systematically, the Sullivan scale was considered as a suitable method for measuring cerumen.

In accordance with the expected result, the level of inter-rater agreement of the video-otoscopy group was higher than that of the hand-held otoscopy group. There were two main possible reasons. Firstly, the video-otoscope provided images with higher quality wherein the larger screen display and the brighter light contributed. The video-otoscope had the image enlarged digitally and been shown on a screen with 450 TV lines horizontal resolution while the hand-held otoscope had the image magnified by a magnifying lens (+2.5 diopeters); the video-otoscope possessed a 150 watt halogen illuminator with a 12 volt (V) power adapter which was powered by the mains while the hand-held otoscope possessed a 3.5V halogen lamp which was powered by two rechargeable nickel-cadmium 3.5 V batteries. Images with higher quality were very likely to have enhanced the diagnostic agreement noted in the video-otoscopy. This is because when the cerumen, tympanic membrane, and/or other ear structures could be clearly identified, the raters would become more competent in fitting the ears' cerumen status to its corresponding scale category. Another research study has documented that high quality images improve the ability to make an accurate otoscopic diagnosis (Mbao, Eikelboom, Atlas & Gallop, 2003), although that study compared the image quality between different types of video-otoscope.

Another reason for the disparities between the otoscopy groups concerned the skills in capturing the otological image. With the ears' images clearly shown on the video-otoscopic screen, the raters would become more capable in altering the viewing angle for a high quality image. In another words, the video-otoscope enabled users to refine their skills in capturing the image, because the screen provided a prompt and concrete feedback.

Apart from the inter-rater agreement, the percentage of agreement among the four scale categories was investigated, and it was observed that the end points of the scale, i.e. 0 and +3, possessed a higher value than the mid points of the scale, i.e. +1 and +2. This may be because the categories 0 and +3 have more precise and concrete definitions. At one end, the category 0 requires a fully visualized tympanic membrane and an almost absent amount of cerumen. At the other end, the category +3 requires a non-visualized tympanic membrane and an occlusive amount of cerumen present. The nature of the cerumen status being classified into these categories is rather discrete. In contrast, the definitions for categories +1 and +2 are less differentiated-both categories yield a non-occlusive amount of cerumen and may yield a partially visualized tympanic membrane. The nature of the cerumen status being classified into these categories is quantity wise different but quality wise similar. These findings exposed a limitation of the Sullivan scale. The mid points of the scale may not be so clearly differentiated as to allow consistent classification of the amount of cerumen, which may account for the fair to moderate level of inter-rater agreement of the scale.

The grand mean percentage of agreement was compared across the two otoscopy groups and there were two main findings. First, the difference between the end points and the mid points of the scale was larger in the hand-held otoscopy group than in the video-otoscopy group. This was because the hand-held otoscopy group yielded a particularly low percentage of agreement for the categories +1 and +2 (59% in average). Raters might have found it difficult to quantify the amount of cerumen precisely with the lower quality hand-held otoscopic image. Second, the percentage of agreement was consistently higher in the video-otoscopy group than in the hand-held otoscopy group for each scale category. This finding provided support to the higher level of inter-rater agreement obtained for the video-otoscopy group. As discussed in above, the high quality video-otoscopic image led to a high diagnostic accuracy. Thus a higher percentage of agreement was obtained.

This study attempted to explain the degree of inter-rater agreement in terms of raters' self-rated confidence towards otoscopy. The whole rating procedure had been divided into three sub-processes, the operation of the instrument, the identification of the images of tympanic membrane and cerumen, and the rating by Sullivan Scale. The level of inter-rater agreement was analyzed between the confident group and the less confident group for each sub-process, to determine which process impacted the most on the inter-rater agreement. It was found that there was no positive trend in the relationship between raters' self-confidence and the AC_1 statistic for the hand-held otoscopy group. For video-otoscopy group, some processes (the identification of tympanic membrane and the rating of actual client ears) reported a positive relationship while some processes reported a negative relationship (the operation of the instrument and the rating of photo/picture) between raters' self-confidence and the AC_1 statistic. However, no matter in the hand-held otoscopy group or the video-otoscopy group, the ratings of both confident and less confident raters scattered similarly, which meant that the raters' self-confidence did not affect their actual ratings. Besides, the level of inter-rater agreement varied slightly between the confident group and the less confident group. These findings did not support any relationship between raters' self-confidence and the inter-rater agreement of otoscopy, or at best a weak relationship.

The author takes a conservative view about the absence of relationship between raters' self-confidence and the inter-rater agreement of otoscopy. A research study by Patricoski, Kokesh, Gerguson, Koller, Zwack, Provost, et al. (2003) examined the inter-rater agreement of in-person examination and remote video-otoscope imaging in assessing the tympanic membrane following surgical placement of tympanostomy tubes. In that study, two otolaryngologists served as raters and their self-confidence was documented with respective to each ear, instead of towards all the subjects in general. Their concerned self-confidence level focused on the diagnostic (rating) process only. Result showed that self-confidence was significantly correlated with the inter-rater agreement, as well as with the quality of the video-otoscopic image. This finding differs from the present study's, which might be explained in terms of the number of raters, the amount of clinical experience of the raters, the nature of the viewed ears and their diagnosis, and the methodology used in evaluating rater confidence level. Further study is needed to investigate raters' confidence towards otoscopy and its relationship to inter-rater agreement.

Although raters' self-confidence in this study was unrelated to the level of inter-rater agreement, it provided practical implications for future training workshops. It was observed that the distribution of self-rated confidence in the two otoscopy groups were quite different in several parameters. For example, the proportion of raters who found themselves confident in operating the hand-held otoscope is three-fold to that in using the video-otoscope; the proportion of raters who found themselves confident in rating by Sullivan scale using video-otoscopy is a double of that using the hand-held otoscopy. This information is of importance in shaping training workshops in the future. The training could be targeted more on the areas that raters have been found to have more self-perceived difficulty with.

This research study had highlighted the importance of structured training in the use of otoscopy. It is known that both clinical experience and training combine to formulate an accurate otoscopic diagnosis (Wormold et al., 1995). However, the undergraduates in the present study lacked direct client contacts, thus structured training was particularly crucial to improve their otoscopic skills. After attending the workshop on otoscopy, the undergraduates were able to use a simple rating scale for outer ear status. Many of them reported that the training was practical, and that they acquired the basic otoscopic skills within the session. The high efficiency of structured training has laid a valuable ground work for other otoscopic training programs. An example would be to train the parents of children who have recurrent otitis media to use home otoscopy, which may help to prevent frequent physician visits and overuse of antibiotics (Berberich & Johnston, 2000). Nonetheless, what needs to be stressed is that structure training is vital to the initial acquisition of otoscopic skills, but the skills need to be supported by extensive clinical experience.

The time since training may be an affecting factor for the rating accuracy. This is because skills are best learnt if they are accompanied by immediate and extensive practice. Raters would have become less capable in making an otoscopic diagnosis if the training had not been quickly followed by practice. Thus the training workshop in the present study was presented only three days in advance of the rating session.

This study had provided additional support to some previous research findings. Certain groups of people are more likely to develop impacted cerumen, for example, the elderly (Flugrath, Lutz & Hamdy, 1993; Ney, 1993; as cited in Moore et al., 2002). During the screening process, the author made a gross comparison on the amount of cerumen found in elderly (age 60 or above) and younger adult (age range between 18 and 60) subjects. It was observed that the elderly were subjected to higher risk of cerumen accumulation. The proportion of subjects with an author rating of +2 or above was 43% in the elderly group while only 16% in the younger adult group. The prevalence of cerumen impaction in elderly could be explained by the physical changes accompanying aging: the hairs in the external auditory canal are longer and denser in elderly persons, which may entrap cerumen; the activity level of the secretary gland decreases in the elderly, which makes the cerumen dries, and less easily removed (Hanger & Mulley, 1992; Ruby, 1986; as cited in Moore et al., 2002).

There were several limitations in this study. Firstly, a gold standard was absent. All the ratings were done by undergraduates of the Division of Speech and Hearing Sciences, who were inexperienced in otoscopy. It was difficult to determine whether raters overestimated or underestimated the amount of cerumen. This limitation has implications when interpreting percentage of agreement findings. It was possible that majority of all raters had made a wrong judgment, which would make the percentage of agreement an index of the proportion of an inaccurate judgment. The best solution would be to invite a group of experienced ENT (Ear-Nose-Throat) surgeons to prepare a “gold standard” consensus rating of each ear. Another limitation is the small sample size and the uneven distribution of the subject ears along the Sullivan Scale. The small sample size limits the generalization of this study as only large sample ensures likelihood of an unbiased sample. In addition, the proportion of subjects in category +3 was relatively small in both otoscopy groups. Thus the AC_1 calculated might be based on results primarily obtained from other scale points.

For future research studies it would be useful to further consider if raters' self-confidence affects in the inter-rater reliability of otoscopy, and how self-confidence level and inter-rater reliability interact. Moreover, it is suggested to examine if the level of inter-rater agreement of otoscopy improves when experienced clinicians served as raters. It has been stressed by many authors that clinical experience plays an essential part in an accurate otoscopic diagnosis. Thus the level of inter-rater agreement is hypothesized to be positively related to clinical otoscopic experience.

In conclusion, the Sullivan scale would be a suitable means to document the amount of cerumen. It makes possible an easy-to-use means for communication of external ear status among practitioners. It could be widely used in daily clinical practice since otoscopy is available in most clinics. On the other hand, the Sullivan scale may have more utility if the definitions of the categories +1 and +2 are refined. This is because the mid points of the scale appear less capable of reliably classifying the amount of cerumen present. Finally, video-otoscopy could be regarded as a preferred primary procedure rather than hand-held otoscopy in rating the amount of cerumen, because of its better inter-rater agreement.

References

American Speech-Language-Hearing Association. (2001). *Scope of practice in speech-language pathology*. Rockville, MD: Author.

Berberich, F. R., & Johnston, K. (2000). Antibiotic use and parental home otoscopy. *Pediatrics, 105,* 159-160.

Burkhart, C. N., Kruge, M. A., Burkhart, C. G., & Black, C. (2001). Cerumen composition by flash prytolysis-gas chromatography/mass spectrometry. *Otology & Neurotology, 22,* 715-722.

Crandell, C. C., & Roeser, R. J. (1993). Incidence of excessive/impacted cerumen in individuals with mental retardation: A longitudinal investigation. *American Journal of Mental Retardation, 95,* 568-574.

Guest, J. F., Greener, M. J., Robinson, A. C., & Smith, A. F. (2004). Impacted cerumen: Composition, production, epidemiology and management. *Quarterly Journal of Medicine, 97,* 477-488.

Gwet, K. (2001a). *Handbook of inter-rater reliability: How to estimate the levels of agreement between two or multiple raters.* Gaithersburg, MD: STATAXIS Publishing Company.

Gwet, K. (2001b). *Statistical tables for inter-rater reliability testing: Tables providing exact t critical values for testing statistical significance.* Gaithersburg, MD: STATAXIS

Publishing Company.

Hawke, M. (2002). Update on cerumen and ceruminolytics. *Ear, Nose & Throat Journal, 81(8 Supp. 1),* 23-24.

Ibraimov, A. I. (1991). Brief Communication: Cerumen phenotypes in certain populations of Eurasia and Africa. *American Journal of Physical Anthropology, 84,* 209-211.

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

Lewis-Culinan, C., & Janken, J. (1990). Effect of cerumen removal on the hearing ability of geriatric patients. *Journal of Advanced Nursing, 15,* 594-600.

Mbao, M. N., Eikelboom, R. H., Atlas, M. D., & Gallop, M. A. (2003). Evaluation of video-otoscopes suitable for tele-otology. *Telemedicine Journal and e-Health, 9,* 325-330.

Meador, J. A. (1995). Cerumen impaction in the elderly. *Journal of Gerontological Nursing, 21,* 43-45.

Moore, A. M., Voytas, J., Kowalski, D. & Maddens, M. (2002). Cerumen, hearing, and cognition in the elderly. *Journal of the American College of Medical Directors, 3,* 136-139.

Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., et al. (1990). Quality-of-life changes and hearing impairment: a randomized trial. *Annals of*

Internal Medicine, 113, 188-194.

Olusanya, B. O. (2003). Hearing impairment in children with impacted cerumen. *Annals of Tropical Paediatrics, 23,* 121-128.

Patricoski, C., Kokesh, J., Gerguson, S., Koller, K., Zwack, G., Provost, E, et al. (2003). A comparison of in-person examination and video-otoscope imaging for tympanostomy tube follow-up. *Telemedicine Journal and e-Health, 9,* 311-344.

Sharp, J. F., Wilson, J. A., Ross, L. & Barr-Hamilton, R. M. (1990). Ear wax removal: A survey of current practice. *British Medical Journal, 301(6763),* 1251-1253.

Sullivan, R. F. (1993). How video otoscopy benefits hearing health care practice. *Hearing Instruments, 44(4),* 14-17.

Sullivan, R. F. (1995). Audiological applications of video otoscopy. *Hearing Journal, 48 (8),* 10, 41-48.

Sullivan, R. F. (1997). Video otoscopy in audiological practice. *Journal of the American Academy of Audiology, 8,* 447-467.

Sullivan, R. F. (n. d.). *Video otoscopy-based scale of cerumen accumulation/EAC obstruction.* Retrieved January 10, 2005, from www.rcsullivan.com/www/vocerum.htm

Wormald, P. J., Browning, G. G. & Robinson, K. (1995). Is otoscopy reliable? A structured teaching method to improve otoscopic accuracy in trainees. *Clinical Otolaryngology, 20,* 63-67.

Acknowledgements

Special thanks to Dr. Bradley McPherson for his encouragement, support, guidance, and valuable advice.

Thanks are also given to Caritas District Elderly Center, for her kind provision of venue and arrangement of subjects. Last, but not least, the author would like to thank all raters for their participation and support in this study.

Appendix A

Sullivan scale

0: Cerumen virtually absent or present in a small quantity, insufficient to limit audiological access to the ear for any procedure. The tympanic membrane may be visualized fully, including the annulus. Wax removal not indicated.

+1: Non-occlusive minor amount present. Tympanic membrane may be visualized fully. Due to location or texture, cerumen may interfere with probe tube real ear measurement, insert earphone use, hearing aid use or impression-taking. Removal is optional.

+2: Non-occlusive moderate amount present. Tympanic membrane may be partially obscured. Cerumen is likely to interfere with probe tube and ear insert measurements as well as hearing aid use. Likely to distort ear impression-taking with the possibility of wax impaction as a consequence. Removal is advised.

+3: Occlusive amount present obstructing fully a view of the tympanic membrane. Likely to interfere with ALL audiological procedures including earphone and sound field hearing testing. May add a conductive overlay. Obstruction of the hearing aid receiver and vent is likely, reducing effective gain and output. Removal is essential.

Appendix B

Workshop Outline

Anatomy of the Ear

1. Outer/Middle/Inner Ear
2. Outer Ear – Pinna, Ear Canal
3. Outer Ear – Tympanic Membrane

Disorders of the Outer Ear

1. Pinna – Deformities/Skin Disorders
2. Ear Canal – Exostoses/Otitis Externa/Foreign Bodies
3. Tympanic Membrane – Otitis Media/Perforation
4. The Commonest Disorder – Cerumen

Otoscopy

1. Why Use an Otoscope?
2. Types of Otoscopes
3. Using an Otoscope

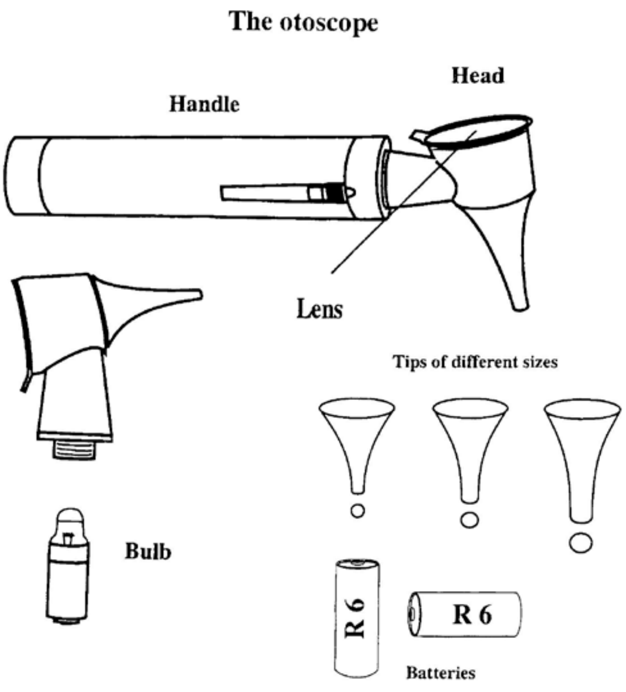
The Sullivan Scale

1. Why Rate Cerumen?
2. How to Use the Sullivan Scale

Practical Session

1. Simulation Model Rating
2. Real-life Rating

Q & A Session



Appendix C
Consent Form

香港大學言語及聽覺科學部

以"Sullivan scale"量度耳垢累積程度之可靠性之 研究

同意書

本部將進行一項研究，目的是提供數據證明"Sullivan scale"量度耳垢累積程度之可靠性。此項研究由本部副教授 Dr. Bradley McPherson 帶領研究員陳軒悠同學進行。

現誠邀閣下在二零零五年一月十七日，往明愛中區長者中心參與研究。閣下之右耳將會由本部副教授 Dr. Bradley McPherson、研究員陳軒悠同學及十五至二十位言語及聽覺科學部之學生以手動式或攝錄式檢耳鏡量度耳垢累積之程度，需時約 30 至 45 分鐘。

是次研究，研究員陳軒悠同學將通過抽籤，安排閣下接受手動式/攝錄式檢耳鏡之量度。是項研究將有助閣下了解耳道及耳膜的情況，對閣下並沒有害處。而閣下的支持，將有助於我們了解"Sullivan scale"的可靠性及比較那一款式之檢耳鏡能夠更加可靠地量度耳垢累積。

所有資料只會作是次研究之用，並予以保密。

我們十分感謝閣下的的支持及參與。如有任何疑問，請致電 9616 4706 與陳軒悠同學聯絡。

****

本人 ____________________ 同意參與是項研究。玆證實上述所有事項，研究員已向本人詳細解釋，本人亦完全明白一切有關安排。

參加者簽署：	研究員簽署：
____________________________________	____________________________________
參加者姓名: 聯絡電話：	陳軒悠
日期：______________________________	日期：______________________________

Appendix D

Recording sheet

Rater's name: ________________ Date of investigation: ________________

Please rate the amount of cerumen accumulation according to the Sullivan scale and mark the result with respect to the patient's number. (Please use a "O" Circle to indicate the rating)

Types of Otoscope	**Patient's number**	**Sullivan scale**			
Hand-held Otoscope	O1	0	+1	+2	+3
	O2	0	+1	+2	+3
	O3	0	+1	+2	+3
	O4	0	+1	+2	+3
	O5	0	+1	+2	+3
	O6	0	+1	+2	+3
	O7	0	+1	+2	+3
	O8	0	+1	+2	+3
	O9	0	+1	+2	+3
	O10	0	+1	+2	+3
	O11	0	+1	+2	+3
	O12	0	+1	+2	+3
	O13	0	+1	+2	+3
	O14	0	+1	+2	+3
	O15	0	+1	+2	+3
	O16	0	+1	+2	+3
Video-otoscope	V1	0	+1	+2	+3
	V2	0	+1	+2	+3
	V3	0	+1	+2	+3
	V4	0	+1	+2	+3
	V5	0	+1	+2	+3
	V6	0	+1	+2	+3
	V7	0	+1	+2	+3
	V8	0	+1	+2	+3
	V9	0	+1	+2	+3
	V10	0	+1	+2	+3
	V11	0	+1	+2	+3
	V12	0	+1	+2	+3
	V13	0	+1	+2	+3
	V14	0	+1	+2	+3
	V15	0	+1	+2	+3
	V16	0	+1	+2	+3

Appendix E

Schedule of the rating program

(R-raters, O-subject ears in hand-held otoscopy group, V-subject ears in video-otoscopy group)

<table>
<tr><td rowspan="2">9:30-9:55</td><td>Patients</td><td>Hand-held Otoscopy (O1)</td><td>Video-otoscopy (V1)</td></tr>
<tr><td>Raters (R)</td><td>R1
R2
:
:
R16
R17</td><td>R1 - R17</td></tr>
<tr><td rowspan="2">9:55-10:20</td><td>Patients</td><td>Hand-held Otoscopy (O2)</td><td>Video-otoscopy (V2)</td></tr>
<tr><td>Raters (R)</td><td>R2
R3
:
:
R17
R1</td><td>R1 - R17</td></tr>
<tr><td colspan="4">:
LUNCH BREAK
:</td></tr>
<tr><td rowspan="2">16:10-16:35</td><td>Patients</td><td>Hand-held Otoscopy (O15)</td><td>Video-otoscopy (V15)</td></tr>
<tr><td>Raters (R)</td><td>R15
R16
:
:
R13
R14</td><td>R1 - R17</td></tr>
<tr><td rowspan="2">16:35-17:00</td><td>Patients</td><td>Hand-held Otoscopy (O16)</td><td>Video-otoscopy (V16)</td></tr>
<tr><td>Raters (R)</td><td>R16
R17
:
:
R14
R15</td><td>R1 - R17</td></tr>
</table>

Appendix F

Research on Otoscopy
Questionnaire

Rater's name:

Please rate how confident you are in (please put a "X" to indicate your choice)

	1 Not confident at all	2	3 Some-what confident	4	5 Very confident
1. using hand-held otoscope after the training workshop	1	2	3	4	5
2. using video-otoscope after the training workshop	1	2	3	4	5
3. identifying the tympanic membrane by using the hand-held otoscope	1	2	3	4	5
4. identifying the tympanic membrane by using the video-otoscope	1	2	3	4	5
5. identifying cerumen by using the hand-held otoscope	1	2	3	4	5
6. identifying cerumen by using the video-otoscope	1	2	3	4	5
7. rating the amount of cerumen by Sullivan Scale (photo/picture)	1	2	3	4	5
8. rating the amount of cerumen by Sullivan Scale with a hand-held otoscope (real ear)	1	2	3	4	5
9. rating the amount of cerumen by Sullivan Scale with a video-otoscope (real ear)	1	2	3	4	5

Appendix G

Raters' Confidence Level towards Otoscopy

Please rate how confident you are

Self-rated confidence	Less confident	1	2	3	4	5	Confident	Total
Q1/Q2 In using hand-held otoscope / video-otoscope after the training workshop								
Hand-held otoscopy	4 (24%)	0	0	4	12	1	13 (76%)	17
Video-otoscopy	13 (76%)	0	3	10	3	1	4 (24%)	17
Q3/Q4 In identifying the tympanic membrane								
Hand-held otoscopy	8 (47%)	0	1	7	6	3	9 (53%)	17
Video-otoscopy	6 (35%)	1	0	5	10	1	11 (65%)	17
Q5/Q6 In identifying cerumen								
Hand-held otoscopy	4 (24%)	0	0	4	7	6	13 (76%)	17
Video-otoscopy	1 (6%)	0	1	0	12	4	16 (94%)	17
Q7 In rating the amount of cerumen by Sullivan Scale (photo/picture)								
	8 (47%)	0	0	8	8	1	9 (53%)	17
Q8/Q9 In rating the amount of cerumen by Sullivan Scale (real ear)								
Hand-held otoscopy	12 (71%)	0	2	10	5	0	5 (29%)	17
Video-otoscopy	8 (47%)	0	2	6	9	0	9 (53%)	17
Total	64 (42%)	1	9	54	72	17	89 (58%)	153