The Development of Speech Perception Tests for Children in the Indonesian Language

Dahlia Sartika

M.D. (Universitas Kristen Indonesia & Universitas Indonesia)

Grad.Dip.Aud. (University of Melbourne)

M.Audiology (Macquarie University)

Audiology
Department of Linguistics
Faculty of Human Sciences
Macquarie University
Sydney

Research project submitted as a requirement for Doctor of Philosophy

December 2014

DEDICATION

This dissertation is lovingly dedicated to the Indonesian children whom I highly hope would gain benefit from this study.

Every challenging work needs self-efforts as well as guidance of elders. My humble effort I dedicate to my loving mother, late father and my auntie (Soesmalijah Soewondo) who is also like my second mother.

TABLE OF CONTENTS

List of Tables

Chapter 5

Chapter 7

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 4

Chapter 7

List of Appendices

ABSTRACT

Well-designed audiometric speech tests for Indonesian children are not currently available. This study involved the development of the Indonesian-Speech Recognition Threshold Test (INDO-SPRITT) for children with mild to moderately – severe hearing loss, which measures the speech recognition thresholds and can be used to: (i) complement pure tone audiometry; (ii) provide unaided speech thresholds and to cross-check pure tone audiometry results; (iii) demonstrate the impact of hearing loss on speech perception to parents and teachers; (iv) provide aided speech thresholds for patients with hearing aids or cochlear implants; and (v) provide a method of demonstrating aided advantage to parents and teachers.

One hundred and fifteen Indonesian children with normal hearing and 16 children with severe to profound hearing loss participated. Results showed that INDO-SPRITT materials were familiar to normal hearing children as young as 4 years and 6 months to severe to profound hearing-impaired children from the age of 7 years and 6 months. Results from the assessment of speech recognition thresholds (SRT) with children who had hearing ≤ 20 dBHL indicated that INDO-SPRITT was a reliable and valid test. The mean SRT in this study provides a normative reference against which the SRT of hearing-impaired individuals can be compared.

As the majority of hearing impaired children in hearing institutions are severely to profoundly deaf, a worthwhile outcome of the study on INDO-SPRITT led to the development of the Indonesian Speech Perception Assessment for Severe to Profound Hearing Loss (INDO-SPASP). This assessment distinguishes between children who can perceive the spectral components of speech and children who can perceive only the time and intensity patterns of speech. Preliminary results for seven children (8-13 years) with severe to profound hearing loss indicate that: (i) INDO-SPASP can be administered successfully to Indonesian children; and (ii) children with hearing loss over 90 dB identify words primarily on the basis of their time and intensity patterns. It is expected that the results from this assessment technique can be used to assist in educational placement, to establish the objectives of auditory training, to measure the effects of auditory training, and to assist in selecting a listening device for the child.

STATEMENT OF CANDIDATE

I certify that the work in this thesis entitled 'The Development of Speech Tests for Children in the Indonesian Language' has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis. The research presented in this thesis was approved by Macquarie University Ethics Review Committee, reference number: HE29APR2005-D0407.

Dahlia Sartika
December 2014

ACKNOWLEDGEMENTS

I take this opportunity to express my gratitude to everyone who supported me throughout the course of this PhD project. It would not have been possible to complete this project without the kind support and help of many individuals and organizations.

My PhD candidature has been a unique one amongst other PhD candidatures, because each of my supervisors mentioned below has had a role as my principal supervisor at different stages. My current principal supervisor is Dr Mridula Sharma, preceded by Dr Robert Mannell who as associate supervisor and Prof. Philip Newall, the adjunct supervisor preceded by Dr Robert Mannell. I would like to express my profound gratitude and deep regard to all of them for their guidance and invaluably constructive criticism, as well as for providing the necessary information required for the project.

Mridula, I also would like to sincerely thank you further for taking me on as your PhD student at the last stage of my candidature, which may have been difficult to understand and grasp the whole concept of my research and to understand me as an individual.

Robert, I also would like to appreciate you further for your time and help for the last few months despite your health problem. Last but not the least, further profound and heartfelt appreciation to Prof. Philip Newall for his kind personal support throughout my PhD candidature.

Furthermore, I am highly indebted to Associate Prof. Norm Erber, Prof. Harvey Dillon and Dr Robyn Cantle Moore for giving me the attention, time and inspirational guidance, invaluably knowledge and friendly advice necessary to help me complete this task.

My special gratitude and appreciations to Chris Callaghan, Dr Siti Faisa Abiretno, Peter Bosloper, Roger Lovegrove, Prof. Soewito and all those who provided me with the facilities that were required for the PhD project. I am also sincerely grateful to Associate Prof. David Butt, Associate Prof. Mehdi Riazi and Dr Stephen Moore for their kind assistance and wisdom.

I would like to extend my sense of obligation to Renwick Centre Royal Institute for Deaf and Blind Children, Carlingford Rotary International Club, Paxton Barrand Hearing Clinic, Karnamanohara Deaf School and PT Kasoem for their valuable contribution and cooperation.

I am also very grateful and highly appreciative to Prof. Goeran Hammarstrom, Sofia Baig and John Newall, for their precious help and sharing their truthful and illuminating views on a number of issues related to the project.

A profound sincere appreciation to Lei Cameron who was not only instrumental in helping me put all the chapters and pages together to become a professional document in such a short time but she also helped me in other aspects of the project, more than is expected of her.

I would also like to express a deep sense of gratitude to Sylvia Liu, Albert Talapessy, Joanne Jung-wook Hong, Claire Urbach, Robyn Little, Anuradha Sengar and to all my dear friends who helped and supported me during a long period of sick leave from my PhD candidature.

Furthermore, my special thanks also go to Cristy Newall and Hayley Mannell for their kind and friendly gesture.

Last but not the least, I would like to express a tremendous heartfelt gratitude towards my mother, late father, brother, sisters and my extended family for their unconditional love and support, which help me to complete this project. The blessing, help and guidance given by them time to time shall carry me a long way in the journey of life on which I am about to embark.

ABBREVIATIONS

ANT	Auditory Numbers Test
ASHA	American Speech-Language-Hearing Association
BI	Bahasa Indonesia
CAT	Children's Auditory Test
CNC	consonant-nucleus-consonant
dB	decibel
DIP	Discrimination by Identification of Pictures
EARS	Evaluation of Auditory Responses to Speech
E.foil	Effective foil
ESP	Early Speech Perception
ESP	Early Speech Perception
FA	Fletcher average
FAHL	frequency average hearing threshold level
FFT	Fast Fourier Transform
GASP!	Glendonald Auditory Screening Procedure
GBD	Global Burden of Disease
HELOS	Hearing Loss Simulator
HINT	Hearing In Noise Test
INDO-SPASP	Introduction to Indonesian Speech Perception Assessment for Severely-Profoundly Hearing-Impaired children
INDO-SPRITT	Indonesian Speech Recognition Threshold Test
MLV	monitored live voice
MTP	Monosyllabic, Trochee, Polysyllabic
MTS or MonSTr	Monosyllable, Trochee, Spondee
NU-CHIPS	Northwestern University Children's Perception of Speech
PEST	Parameter Estimation by Sequential Testing
P-I	performance-intensity
PLOTT	Plant and Wescott Test

PSI	Paediatric Speech Intelligibility
PTA	pure tone average
SA	speech average
SD	standard deviation
SDT	speech detection threshold
SERT	Sound Effect Recognition Test
SLM	sound level meter
SPAC	Speech Pattern Contrast test
SRT	speech recognition threshold
ST	speech threshold
TACT	Text Analysis Computing Tools
THRIFT	Three-Interval Forced Choiced test of speech pattern contrast perception
UKM	University Kebangsaan Malaysia
WHO	World Health Organisation
WIPI	Word Intelligibility by Picture Identification
WRS	word recognition score
YLD	years of life lived with disability

CHAPTER 1:
BACKGROUND

Deafness is the most common sensory disability in the world. The problem is disproportionately high in the South-East Asia Region where every third deaf person in the world is a South-East Asian. The World Health Organisation (WHO) estimates approximately 38,000 deaf children are born in South-East Asia annually. This would mean that over 100 deaf infants are born in the region daily. Deafness in infancy and childhood makes an immense impact on communication, education, employment, and quality of life. This is influenced by the age of the child when identification took place and the time of onset of the hearing loss." (WHO, 2004).

The population in Indonesia is estimated at 252.8 million (World Population Review, 2014) and, with a prevalence of 2% (which is the figure for the Asia Pacific region quoted by WHO in 2012) this would mean that about five million children in Indonesia have a hearing impairment. The WHO has also estimated that two-thirds of those with severe-profound hearing loss live in developing countries (Kumar, 2001). The majority of hearing-impaired children found in hearing clinics and schools for the deaf in Indonesia are severely or profoundly deaf, most living in poor socio-economic conditions.

Hearing impairment in high-income countries has been shown to cause very large financial costs. For example, in Australia, the real financial costs of hearing loss was $11.75 billion in 2005, however, does not take into account the net cost of the loss of well-being (disease burden) associated with hearing loss, which is a further $11.3 billion (Access Economics, 2006). This situation is also likely to be the case in middle- and low-income countries. Hearing impairment is a cause and consequence of poverty, especially in middle- and low-income countries such as Indonesia. Little data is available to document the impact on the poor, but it is evident that poor persons who also have hearing loss are more disadvantaged and may be called the 'poorest of the poor' (Smith, 2008).

Furthermore, as in many developing countries, Indonesia lacks resources for diagnosis and rehabilitation for those with hearing impairment/deafness. The output of available human resources, such as teachers of the deaf, speech therapists and audiometricians are less than optimal because of a variety of reasons attributable to training, deployment and work environment (WHO, 2004).

1.1 Consequences of hearing loss in children

The most important consequence of hearing loss in children is delayed speech and language development, which can lead to social and emotional problems, as well as academic failure. The longer a child's hearing loss remains undetected and untreated, the worse the outcome is likely to be (Northern & Downs, 2002; Yoshinaga-Itano, 2003b). In addition to individual effects, hearing loss makes a large contribution to the burden of disease and it substantially affects social and economic development in communities and countries. Childhood deafness accounts for significant years of life lived with disability (YLD). WHO used the Global Burden of Disease (GBD) to assess the impact of an illness or injury and hearing impairment is the second leading cause of YLD and is the 15th leading contributor to GBD (WHO, 2005).

Hearing loss is often undetected in all age groups in developing countries because routine hearing screening is uncommon, especially for children (McPherson & Olusanya, 2008; Leigh et al., 2010). In Indonesia, as in Europe many years ago (Branson & Miller, 2002), a child's failure to talk or understand speech at a normal age is often recorded as 'retarded'. When a child's deafness is finally revealed, the child is often referred to as 'deaf and dumb'. It is common for deafness to be ignored in infancy, even by intelligent and observant parents, and it is only diagnosed belatedly when the child fails to develop speech. The additional diagnosis of 'dumb' or 'retarded is also widespread and may remain throughout the individual's life.

The commonly known term 'deaf' is so general that it constitutes vastly differing degrees of deafness, ranging from mildly to profoundly deaf. Some hearing institutions, however, prefer to use terms 'hearing loss' 'hard of hearing' or 'hearing impairment' which also ranges from mild to profound. A person with profound hearing loss has virtually no ability to form speech or understand it, because listening is the sole means of receiving the necessary information to develop it, unless sign language is used.

It is not merely the degree of deafness that matters but crucially the age or stage at which it occurs. The age at which hearing loss occurs is crucial for acquiring a spoken language (Carney & Moeller, 1998; Yoshinaga-Itano, 2003a).

When hearing loss or deafness occurs after the acquisition of speech and language, usually after the age of six, it is referred to as 'post-lingual deafness'. It may develop due to disease, trauma, or as a side effect of medication. Typically, hearing loss is gradual and is often detected by family and friends of affected individuals long before they acknowledge their disability (Cowie & Douglas-Cowie, 1992).

Hearing loss occurring prior to the acquisition of language as a result of a congenital condition or through hearing loss in early infancy is referred as 'pre-lingual deafness' (McPherson & Olusanya, 2008). The period from birth to five years of age is considered the critical/sensitive phase for the development of language, while hearing during the first year of life is crucial for normal acquisition of language (Carney & Moeller, 1997; Yoshinaga-Itano, 2003a). Studies have demonstrated that detecting pre-lingual hearing loss and providing amplification within a family-oriented intervention program in the first year of life are associated with favourable outcomes in speech and language development (Moeller, 2000; Kennedy et al., 2006; Ching, Leigh & Dillon, 2013; Moeller et al., 2013). Pre-lingual deafness impairs an individual's ability to acquire a spoken language. Because babies learn language by listening to the people around them, an undetected or untreated hearing loss can have a devastating effect on a child's ability to develop speech and language skills. A child who cannot hear sounds or differences between sounds will have difficulty understanding words and speech (Dobie & Hemmel, 2004).

When hearing is absent at birth or lost in infancy before the language is acquired, its impact on the quality of life is often hard to understand by those with normal hearing (and even by the post-lingually deafened). For people who have never heard or who have no possible auditory memories, images, or associations, there can never be even the illusion of sound. They live in a world of utter silence. They are the so-called 'pre-lingual profoundly deaf' (sometimes referred to as 'stone deaf' by laypeople) (Wright, 1990).

The predicament of pre-lingual deaf children is unique and people tend to think deafness is less serious than blindness, to see it as a disadvantage, or nuisance, or a handicap, but rarely as devastating in a radical sense (Sacks, 1991; Branson & Miller, 2002). Deafness, unlike blindness, is an unseen disability. The general public is, therefore, largely ignorant of the phenomenon of hearing loss (WHO, 2004).

Pre-lingually deaf children, who are unable to hear their parents, risk being severely retarded, if not permanently defective, in their grasp of language unless early and effective measures are taken. To be defective in language is one of the most desperate calamities for anyone because it is only through language that we enter into our human estate and culture, communicating freely with our fellows, acquiring and sharing information. If we cannot do this, we will be bizarrely disabled and cut off from our surroundings, desires, endeavours and native capacities. And indeed, we may never be able to realise our intellectual capacities because we appear mentally defective (Sacks, 1991; Branson & Miller, 2002; Yoshinaga-Itano, 2003b). It is for this reason that the pre-lingual profoundly deaf are often considered less intelligent, resulting in the term 'deaf and dumb'. They are regarded by an unenlightened law as 'incompetent' to inherit property, to marry, to receive education and to have adequately challenging work. Basically, they are being denied fundamental human rights. It was not until the 18th century when steps were taken to remedy this situation in developed countries, which resulted in the perception and situation of deaf people being radically altered (Lane, 1984, 2006).

Has the situation in Indonesia been rectified? Perhaps so for the very few people who are fortunate enough to possess hearing devices, or who have access to special education, but, not for those who live under poor socio-economic conditions, especially profoundly deaf children in Indonesia. McPherson (2008) also indicated that the cost associated with diagnostic and rehabilitative audiological services in developing countries may be a barrier to all except the privileged few.

With regard to the pre-lingual profoundly deaf population in Indonesia, many questions can be asked. Why are uneducated pre-lingual profoundly deaf people isolated in nature and unable to communicate with other people? Why are they reduced to a state, which might be described as imbecility? Is their biological constitution different from ours? Do they have everything they need for having sensations, acquiring ideas and combining them to do everything that we do? Do they receive sensory impressions from objects as we do? Why are pre-lingual deaf people expected to be less intelligent than those who can hear? These questions were addressed in many developed countries several hundred years ago (Lane 1984, 2006), but many of these questions may have not been adequately addressed in developing countries such as Indonesia despite major advances in devices and interventions in other countries (Olusanya, 2005a; Olusanya et al., 2007).

One of the fundamental answers to some of the above questions lies in the use of symbols. The uneducated pre-lingual profoundly deaf person has no symbols for fixing and combining ideas. Therefore, there is a fundamental communication gap between them and other people. This may result in a condition where they are virtually without language, which may have many implications in the area of cognition. We speak to inform other people, as well as to tell ourselves what we think. Speech is part of thought (Geers & Moog, 1989; Sacks, 1991; Blamey, 2003).

It is important to realize that in developing countries such as Indonesia, the poor uneducated languageless deaf may appear to be unintelligent. Their intelligence (although present and perhaps in abundance) is locked up for as long as the lack of language lasts (Sacks, 1991).

How do we unlock a person’s capacities and allow him/her to grow and think? How do we change their predicament? One of the answers lies in the awareness of the magnitude and consequences of deafness or hearing impairment in all parts of society. Awareness will lead to a stronger advocacy and political will to deal with it.

Another answer lies in facilitating the acquisition of language, which in developing countries may involve signing.

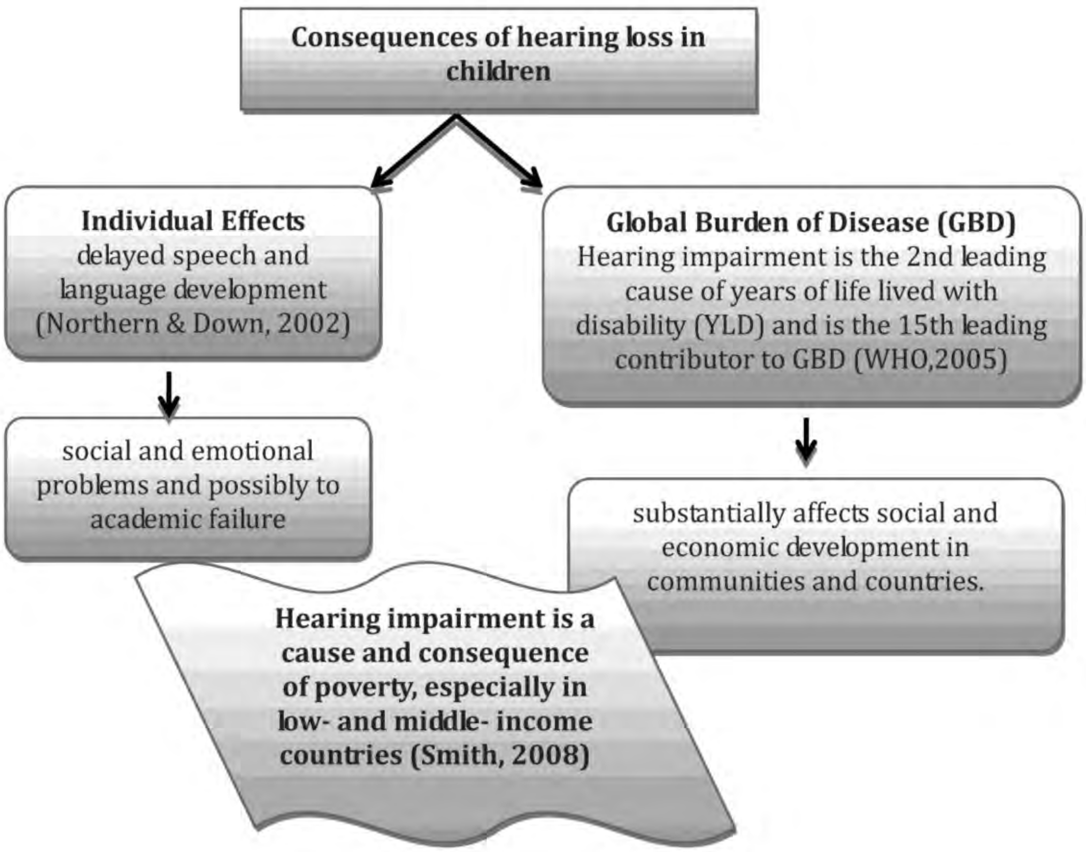


Figure 1.1 Summary of consequences of hearing loss in children

1.2 Awareness of deafness

Despite being the most common sensory disability, deafness has received little attention in health development agendas of many countries including Indonesia. The health care system in the developing world has been traditionally preoccupied and solely assessed with mortality indices (Olusanya, 2005b; Leigh et al., 2010). Thus, it is unusual to find benchmarks for non-fatal conditions such as deafness. The consequence of this is the rapidly increasing burden of deafness. One of the reasons that deafness and hearing impairment received little attention is due to a lack of evidence-based information and awareness of the magnitude and consequences of deafness and hearing impairment in all parts of society. Furthermore, there is a lack of strong advocacy and political will to deal with it, which leads to a lack of resource programs. The collection of accurate data will help to raise awareness among the general public, opinion leaders, decision makers and health planners, as well as help for a stronger advocacy (WHO, 2004).

1.3 Needs in hearing health care

There are four essential needs in healthcare for the hearing impaired: (i) preventive programs; (ii) screening programs; (iii) diagnostic services; and (iv) rehabilitative programs (McPherson, 2008). To assist in meeting some of these needs in the diagnostic and rehabilitation area, two diagnostic tools have been developed in this study to assess the impact of hearing impairment on children's speech perception.

1. INDO-SPRITT measures the speech recognition threshold (SRT) and can be used: (i) to complement pure tone audiometry; (ii) to demonstrate the impact of hearing loss on speech perception to parents and teachers; (iii) to provide an aided speech threshold test for patients with hearing aids or cochlear implants; and (iv) to provide a method of demonstrating aided advantage to parents and teachers.

2. INDO-SPASP provides a measure of speech perception ability in subjects with severe to profound hearing loss and can be used: (i) to assist in educational placement; (ii) to establish the objectives of auditory training; (iii) to measure the effects of auditory training; and (iv) to assist in selecting a listening device for the child.

The collection of data from INDO-SPRITT and INDO-SPASP should help hearing impaired children by increasing the awareness of parents, teachers and the general public of the implications of deafness. The test procedures will help parents and teachers to identify the limited speech perception ability of children with hearing impairment. This may raise awareness of the problems encountered by children with a hearing loss, hopefully leading to stronger advocacy.

CHAPTER 2: INTRODUCTION TO INDONESIAN SPEECH RECOGNITION THRESHOLD TEST (INDO-SPRITT)

Auditory assessment using speech stimuli is referred to as speech audiometry. Speech audiometry involves the assessment of sensitivity (audibility component) for speech as well as assessment of clarity (distortion component) when speech is heard (Plomp, 1978). The audibility component is quantified through assessment of speech thresholds. There are two main types of speech threshold (ST) measures: speech detection threshold (SDT) (for detecting speech) and speech recognition threshold (SRT) (the level at which it is possible to understand a given percentage of speech). SRT is the most frequently measured speech threshold. The distortion component is typically quantified through assessment of speech at suprathreshold level which is commonly known as word recognition score (WRS) (McArdle & Hnath-Chisolm, 2009; Lawson & Peterson, 2011).

Plomp (1978, 1986), pointed out that communication handicap caused by hearing impairment can be defined quantitatively as the elevation of SRT in quiet and/or noise over that of a normally-hearing individual. Elevation of the SRT is caused by the combined effects of two components in his model, audibility, which only affects the SRT in quiet, and distortion, which affect the SRT in quiet and in noise. The effect of audibility on communication handicap is addressed by amplification, while the effects of distortion is addressed (to some extent at least) by improvement of the S/N ratio.

Standardised speech audiometry materials for children in the Indonesian language or Bahasa Indonesia (BI) have not been developed, though, Soewito et al. in 1984, developed paediatric speech test materials for screening school age children. The word lists were presented by means of live voice at whispered voice level and in an open set condition. The children were asked to repeat the word heard. Whilst this was a useful test procedure, this test format was not suitable for use with younger children. Furthermore, oral response was inappropriate for those children with articulation problems because the tester could not be sure whether such a response was due to faulty hearing or faulty articulation, or both (Markides, 1987). Therefore, it seemed reasonable to develop a standardised speech test material for children with a more appropriate response mode.

On the other hand, there are many standardised paediatric speech tests in English currently available in the world, such as PBK-50 word lists by Haskins, 1949; Word Intelligibility by Picture Identification (WIPI) by Ross and Lermann, 1970; Discrimination by Identification of Pictures (DIP) by Siegenthaler and Haspiel, 1966; Northwestern University Children's Perception of Speech (NU-CHIPS) by Katz and Elliot, 1978; newer Paediatric Speech Intelligibility (PSI) test by Jerger et al., 1980; Kendall Toy Test by Kendall, 1954; Sound Effect Recognition Test (SERT) by Finitzo-Hierber et al., 1980; Manchester Junior Lists by Watson, 1957; AB Isophonomemic Word Lists by Boothroyd, 1968 and BKB Sentence Lists by Bench and Koval, 1979.

Paediatric speech tests in non-English languages are also available in such tests in an Aboriginal Language by Plant (1990), a Cantonese Speech Audiometry test for Children of Hong Kong by Kei et al. (1991) and Saudi Arabic speech audiometry for children by Ashoor & Prochazka (1985). The principles for the development of a paediatric speech test in a non-English language are basically the same as in English. The main difference is in the construction of the test material that is influenced by the language constraints.

Some principles in the development of speech audiometry for children are different than they are for adults such as: (1) extra-auditory factors: cognitive, motoric and attentional demands of the test should be age-appropriate; (2) the task (e.g., response task) must be interesting and motivating; (3) vocabulary restriction or performance should be independent of vocabulary knowledge and higher-level language ability; (4) the test should not require phonological knowledge or speech production skills (Jerger, 1984; Kirk et al., 1997; Kosky & Boothroyd, 2003).

In addition to these principles, considerations should be made for inconsistencies in a child's alertness, motivation and fatigue (Mendel, 2008).

Furthermore, as in any measurement used to assess one's behavioural performance, speech perception tests for children must be developed with proper attention to sensitivity, validity and reliability issues, and that appropriate rules of test construction and standardisation should be followed. Carefully controlled test methodologies should be used for sensitive and accurate assessments (Mendel, 2008).

In Section 2, psychometric considerations (sensitivity, reliability, validity) (Section 2.1), methodological considerations to meet those principles (Section 2.2), the aims of INDO-SPRITT (Section 2.3) and the research question will be described (Section 2.4).

2.1 Psychometric considerations

Mackersie (2002) noted that an ideal speech perception test should be reliable, highly *sensitive* to differences between test conditions and should correlate well with speech perception abilities in the real world, that is, will have high *validity*. These three concepts (sensitivity, reliability and validity), should be a significant consideration during the selection of appropriate speech material.

2.1.1 Sensitivity

Sensitivity of a speech perception test refers to 'how objectively and accurately a test measures aspects of listeners' speech perception abilities as a reflection of their performance in realistic listening situations' (Mendel & Danhauer, 1997). High sensitivity is present if a small change in the experimental conditions causes a large and repeatable change in test scores (Dillon, 1983). Sensitivity in speech perception test material is maximised if all the items in every list are of equal difficulty (Dillon, 1983) and the lists themselves are of equal difficulty (Bamford & Wilson, 1979).

In addition to the aspects of speech perception tests that ensure sensitivity described above, Bilger (1984) and Elkin (1984) indicated that a sensitive speech perception test also needs to have the following aspects clearly established: Its purpose, the validity (predictive, content, construct and face), reliability, including measurement of the standard error and typical participant variance for the population of interest, reliable and equivalent alternative test forms, and lastly, a clear procedure for test administration, scoring and interpretation.

2.1.2 Reliability

One of the greatest problems in the development of a speech perception test is to determine the reliability of the test being constructed. According to Ruscetta et al. (2005), a test is not useful in research or for documenting a clinical intervention unless reliability is quantified. Issues of reliability are important considerations, both in choosing test material and interpreting test results (Cacae & McFarland, 1998). Reliability refers to the extent to which test results are repeatable at different points in time. Reliability is quantified by taking several measurements on the same participants. Poor reliability will compromise test sensitivity.

Mackie and Dermody stated:

> Reliability refers to the reliance (i.e., confidence) one can place in a measuring device; the certainty one has that the result is consistent with that of other tests used to measure the same thing: the stability of the test itself from one occasion to the next; and the sensitivity of the test to changes in the behaviour is being assessed. (1982, p. 70)

There are four major methods to determine the reliability of tests of speech intelligibility: (i) test re-test reliability; (ii) alternate forms method (inter-list equivalence); (iii) split- half method; and (iv) inter-item consistency (Mackie & Dermody, 1982; Burns, 1994).

Reliability can also be expressed in terms of standard error of measurement. If a listener is given the same test several times, the score obtained from an average of the test scores would approach some value (i.e., the true score) more and more closely. The degree to which a single test score approximates the true score, determines the reliability of the test (Mendel & Danhauer, 1997).

However, a high degree of this test re-test reliability does not necessarily mean that the test has a high validity. That is, even though a measure may be repeatable or precise, the measure itself may not be a valid or accurate measure of the behaviour of interest. Thus, a measure can be reliable but not valid, however, in order for a measure to be valid, it ought to be reliable (Nunnaly, 1978).

2.1.3 Validity

As with reliability, test validity can be assessed in a number of different ways. These include construct validity, content, predictive and face validity (Mendel, 2008; Walden, 1984).

2.1.3.1 Construct validity

Construct validity of a speech perception test refers to the actual ability of the test to measure the abstract construct of speech perception. The construct of the test material should therefore reflect the speech it is being compared to. An example of poor construct validity would for instance be, if nonsense syllables are used to draw conclusions about speech perception in daily communication (Mendel & Danhauer, 1997).

From the research point of view, the most essential form of validity is construct validity (Burns, 1994). It really brings into question what property is in fact being assessed. A speech test would be considered to have construct validity if it provides information on speech perception in communication and the circumstances influencing it (Mackie & Dermody, 1982).

However, speech perception is a complex abstract construct and it is necessary to use a number of speech perception tests to analyse the many parameters, which are present. Thus, a test battery is needed to measure specific aspects of speech recognition such as phoneme recognition, word recognition, sentences spoken by different talkers and various conditions. The results of this battery should then provide an individual’s abilities in daily communication (Mendel, 2008; Walden, 1984). Unfortunately, most clinical situations do not allow sufficient time for the use of such a battery.

It is intended that a test battery will eventually be developed to complement INDO-SPRITT. This important issue is dealt with in section 4.2.4.

2.1.3.2 Content validity

Content validity refers to the relationship between the behaviour studied and the content of the test. With regards to speech perception test development this can be achieved by establishing the criteria that is required for the speech material such as lists that are phonetically balanced, lists of equal difficulty and the use familiar vocabulary. During speech perception test development, it is necessary to use a native speaker of the language in which the test is developed.

Content validity is established during the developmental stages, by determining the extent to which the test provides a representative of the type of test item of interest and the sample of the test items is appropriate for the purpose of the test (Walden, 1984).

2.1.3.3 Predictive validity

Predictive validity refers to the ability of a test instrument to estimate a particular type of behaviour. In speech perception, predictive validity may refer to the ability of a specific test to predict one's speech perception performance in a realistic listening environment on the basis on one's scores on the test instrument. For example, WRS may be used to predict a patient's performance in a realistic communication environment with particular hearing aid.

Predictive validity is also known as the criterion validity of a test and is determined by measuring how well the test correlates with some outside validating criterion (Mendel & Danhauer, 1997). For example, SRT in quiet can be used to predict pure tone threshold, because there is generally good correlation between speech and pure tone thresholds both for people with normal hearing (Fletcher, 1929) and those with hearing loss (Hughson & Thompson, 1942).

2.1.3.4 Face validity

Refers to the extent to which a test instrument appears to measure what it is supposed to measure (Nunnally, 1978; Walden, 1984; Mendel, 2008). The judgment of face validity is determined in the initial stages of test construction to offer a global decision as to whether the plan of content validity is appropriate (Walden, 1984). Thus, an instrument that has good face validity will also have a good content validity.

Hearing tests using speech stimuli have higher face validity than those using pure tones, because communication depends on listening to speech rather than pure tones (Kruger

& Mazor, 1987). Furthermore, children pay closer attention to verbal stimuli such as speech than to non-verbal stimuli such as pure tones (Hardy & Bordley, 1951).

2.2 Methodological considerations

2.2.1 Selection of test vocabulary

2.2.1.1 Context

There are several types of material such as nonsense syllables, monosyllabic words, disyllabic words, polysyllabic words and sentences that can be used for speech discrimination testing.

The context of the test stimuli has been shown to affect the intelligibility of speech. In general, the intelligibility of speech material falls along a continuum of difficulty based on the meaningful information in the utterance. The more information there is the steeper its performance-intensity (P-I) function. Four syllable words are more intelligible than three-syllable words. Three syllable words are more intelligible than two syllable words, and so on. Monosyllabic words are more intelligible than nonsense syllables, and sentences are more intelligible than polysyllabic words (Egan, 1948). A P-I function plotted for single syllable words will become steeper when the same words are heard in sentences (Miller, Heise, & Lichten, 1951). For isolated words it may be that, with longer words, each syllable provides a context which restricts the possible alternatives and therefore lowers the threshold of the word by supplying enough context to exclude most of the errors that would occur for each syllable separately (Cole & Jakimik, 1980).

The findings of the effect of language and context of the intelligibility of speech have important clinical implications. Speech recognition tests are aimed at evaluating the ability of a listener to identify a speech message. When listening to a speech signal in everyday life, the listener not only processes the incoming acoustic properties of the utterance but he/she is also influenced by language redundancy. Therefore, in selecting a test item, it is preferable to use an item that represents a balance of information between acoustic properties and language redundancy.

In this case, in English, monosyllabic words are found to be representative for this kind of situation, as they have minimal lexical and phonetic context, thus they do not have too much redundancy. Thus, this structure will lead to a result which is least contaminated with the effects of contextual variables (Mackie & Dermody, 1982). Many of WRS tests in English use monosyllabic words, which have the advantage of quicker to administer compared to using sentences.

However, unlike English, most of the words in BI are disyllabic or polysyllabic. Therefore, monosyllabic words are inappropriate for use in BI speech tests. Otherwise, using these words will only evaluate the listener's ability to detect the acoustic properties of the speech signal as most of the monosyllables in BI are nonsense syllables.

While many WRS tests in English use monosyllabic words, the preferred materials for the measurement of SRT in the United States are spondaic words. Although, in theory, almost any materials could be used. Spondees (bisyllabic words with equal stress on the two) have the advantage of being homogenous with respect to audibility, or both syllables are just audible at about the same speech intensity level (Stach, 1998), therefore, auditory cues from each syllables are equally available to the ear (Hudgins, et. al, 1946). It was also found that due to the steep slope of the psychometric function of the spondees, the accuracy of the SRT value obtained is better than for test words with less steep slope (Arlinger, 1991). This helps greatly in establishing a threshold for speech. By presenting a series of spondaic words and systematically varying the intensity, one can determine the lowest level at which the individual can identify about 50% of the test items.

However, the use of spondee words may not always be applicable in other dialects or languages. For example, the number of spondee words available in Australian English is somewhat limited and the words are often unfamiliar to young children. As a result of this, monosyllabic words are used for their SRT procedure (Golding & Birtles, 2001). It is interesting to note that in the Aboriginal languages (e.g., Tiwi and Walpiri languages), monosyllabic words are relatively rare. The SRT test which has been developed in the Tiwi language uses trisyllabic words because a suitable set of bisyllabic words could not be obtained, whereas, in the Walpiri language, bisyllabic words with stress on the first syllable of a word are used for its SRT test (due to the rules pertaining to syllable and word formation in the Walpiri language) (Plant, 1990). Furthermore, sentence materials have also been used to measure SRTs such as in the Hearing In Noise Test (HINT) (Nilsson et al., 1994).

As mentioned above, most of the words in Indonesian language are disyllabic or polysyllabic. However, the Indonesian language does not have spondaic or trochaic words because stress is essentially free and communicatively irrelevant. A further discussion about stress patterns in the Indonesian language is to be found in Chapter 5.

There are more than 700 kinds of ethnic languages in Indonesia (Lewis, 2009). In this study, however, the national language which is Bahasa Indonesia (BI) is used instead of using an ethnic language. BI is the language of government and the medium of instruction in schools, and it is used in an increasingly wide sphere of social interaction, including interethnic communication, religion, and mass communication. There is an increasingly large population of speakers for whom Indonesian is their first language. An estimated 23 million people speak BI as a first language and an additional 140 million speak it as a second language (Grimes, 1996; Gordon, 2005). Most children in big cities speak BI for daily communication. However, many children in the villages or small towns usually speak their ethnic languages for daily communication. Therefore, it is intended INDO-SPRITT will mainly be used in big cities such as Jakarta, Surabaya, Yogyakarta and other cities where children use BI for their daily communication.

2.2.1.2 Familiarity and word frequency effects

Word familiarity is generally defined by their frequency of occurrence within the language as reported in established word counts (e.g. word counts by Lorg included 4,5 million words from written material). The assumption is made that words with a high frequency of occurrence will be heard more often by the subject in everyday life and will therefore be more familiar to him/her (Mackie & Dermody, 1982). When a word stimulus is presented to a subject, it will evoke every word in a person’s vocabulary as a possible response. Next, the subject must compare and choose the word that best matches the stimulus, among the possible words in his/her stock of words. Thus, the more common the word is in the language the more readily it is identified, and therefore the word is easier to understand.

Word familiarity is therefore an important aspect in speech testing. The words that are used should be very familiar to lessen the effect of education and intelligence on a listener's performance (McFarlane, 1940). It has been shown that the use of items that are not within the patient's vocabulary cause spurious low scores and lead to unnecessary testing, extra expenses and misdiagnosis. Therefore, the words used must be within the vocabulary of the target population.

In 1961, Owens further examined the word frequency effect on speech intelligibility using a sample of 4.5 million words from Lorg's word count. Owens used different materials that consisted of monosyllabic word lists of different familiarity (unfamiliar with frequency of usage between 0 and 100; moderately familiar between 150 and 500; highly familiar between 1000 and 2000, and they were matched phonemically). The lists with greater familiarity, even to a small degree, were markedly more intelligible.

Studies of word familiarity in children's speech perception testing have shown increasing performance with age on speech tests designed for even the youngest children who were tested. Elliot, Clifton and Servi (1983) found that the developmental improvement in performance could be explained by 'word frequency effects' by which they meant that although the children knew and were familiar with the stimulus words, they had less experience with them than the adults had. Because of these frequency effects, young children (under 8 years of age) needed speech to be at a higher intensity level than for older children of adults in order to be understood (Byrne, 1983). Thus, it is not appropriate to assess a child's speech perception threshold using an adult test. The results would suggest that many children had a hearing loss for speech, when in fact their performance was within the normal range for their age.

2.2.1.3 Phonemic balance

The words used in each list should be phonemically balanced if possible. This balance is sometimes difficult to achieve with short word lists and according to Egan (1948), 50 monosyllabic words in a list are the minimum requirement to achieve phonemic balance.

Materials are considered as phonemically balanced, when the different phonemes occur in the test material at relatively the same frequencies as they occur in the language (Egan, 1948; Lyregaard, 1987). The rationale of this concept is the belief that the phoneme is the most relevant part of spoken language for determining the intelligibility of speech, so that having the phonemes as they occur in the language will be a valid reflection of speech in everyday communication (Mackie & Dermody, 1982). Another reason to have a phonemically balanced word list is that when listeners are totally unable to perceive a particular phoneme which occurs infrequently in normal everyday speech, the handicap they experience is not as severe as it would be had the phoneme been a more frequently occurring one (Lyregaard, 1987).

Phonemic (or phonetic) balance may be thought of as a relation between parent population and test material. In fact the same concept is relevant at the next stage, namely between test material and list. For speech audiometry, test lists are regarded as interchangeable if each has the same phonemic balance. Lyregaard (1987) proposed that the term phonemic equalisation is more appropriate (Lyregaard, 1987).

Lehiste and Peterson (1959) used a slightly different approach to the issue of phonemic balance. They described their words as ‘consonant-nucleus-consonants’ (CNC) because they identified the vowel as the ‘syllable nucleus’ in a word. They pointed out that particular speech sounds are affected by other speech sounds preceding and following them, so that precise phonemic balance is difficult. They also devised 10 lists of 50 words each, with a phonemic balance similar to a corpus of 1263 CNC type of monosyllabic words (selected from Thorndike and Lorg’s word count of 1,000,000 words) rather than the balance of phonemes present in the language as a whole and the order of balancing was to have each initial consonant, each vowel and each final consonant appear with the same frequency of occurrence within each list. An example of a speech discrimination test that uses this approach is NU-CHIPS (Elliot & Katz, 1980).

Thus, Lehiste and Petersen (1959) defined phonemic balance differently from Egan (1948) and they pointed out the difference between phonetic and phonemic balancing. Phonetic balancing is when lists in which the sounds within the list occur with the same frequency as they do in a representative sample of English speech. Phonemic balancing is lists in which each initial consonant, each vowel, and each final consonant appear with the same frequency of occurrence in the test list.

Martin, Chaplin and Perez (2000), however, compared word recognition scores using phonetic balance word lists and similar lists of words that were deliberately not phonetically balanced. They compared scores on subjects with normal hearing and those with sensorineural hearing loss. The scores for the lists were almost identical, which brings into question the belief that the lists are, or need to be phonetically balanced. Dillon and Ching (1995), and Martin, Chaplin and Perez (2000) also pointed out that the relevance of precise fulfilment of phonemically balanced speech perception test material for predicting communicative difficulties in everyday life due to hearing loss is questionable.

Table 2.1 Consonant's of BI (Soderberg & Olson, 2002)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive & affricate	p b		t̪	d	tʃ dʒ		k g	(ʔ)
Nasal	m			n		ɲ	ŋ	
Flap/trill				r				
Fricative		(f)		s (z)	(ʃ)			h
Approximant	w					j		
Lateral approximant				l				

The phonological characteristic of BI comprises of 22 consonant phonemes (table 2.1) , six monophthong vowel phonemes (table 2.2) and three diphthongs, /ai/, /oi/, /au/ (Lapoliwa, 1981; Soderberg & Olson, 2002).

Despite these opinions, an attempt was made to create a phonemically balanced list in INDO-SPRITT and this process followed the term phonemic balance as defined by Egan (1948) and Lyregaard (1987). If the researchers mentioned earlier are correct, this may not be strictly necessary, but it was felt that it was a safer option.

Table 2.2 Vowels of BI

	Unrounded		Rounded
	Front	**Central**	**Back**
High	i		u
Mid	e	schwa	o
Low		a	

BI syllables basically do not have consonant clusters. The words with clusters are usually due to the influence of foreign languages. Most BI syllables consist of one of the following patterns (Dardjowidjojo, 1978):

1. V examples are: i-kan, u-dang.
2. VC examples are: in-dah, um-pan
3. CV examples are: ru-mah, bu-ku
4. CVC examples are: kun-ci, pan-tai

2.2.2 List characteristics

2.2.2.1 List length

The length of test list is determined by the clinical requirements for a test that is quick to administer and has at least the minimum acceptable reliability. Some researchers suggest that a 50-item word list is the minimum length required to meet the requirements for a clinical test (Grubb, 1963). The choice of a 50-word list is usually determined by phonemic balance issues. Others have suggested using shorter test lists to assist in reducing the time required evaluating the hearing of patients in audiological clinics. The shorter lists are obtained either by dividing the 50 word list in two lists of 25 words – the so-called half list (Elpern, 1961; Campanelly, 1962; Resnick, 1962) or by constructing a new list of 25 items but using the same basic word pool as the 50 word lists (Breaky, 1948; Shutts et al., 1964; Campbell, 1965; Margolis & Millin, 1971; Raffin & Thornton, 1977). The shorter lists have been used to lessen the time for testing the elderly, young children and patients who are difficult to test.

Another option is that of Boothroyd (1968) where very short word lists are used. They are not phonemically balanced, but they are isophonemic. Each list consists of the same thirty phonemes to form 10 consonant-vowel-consonant words. In other words, the list has an even distribution of phonemes with each phoneme only occurring once in a list. The advantage of this approach is it allows a more rapid administration to determine a performance/intensity function for patients. When shorter/partial lists are used, one should realise that reliability may be sacrificed, because, the more items which are scored, the more reliable a test is (Dillon, 1983; Dillon & Ching, 1995).

2.2.2.2 Alternative forms of test and list equivalence

In cases where individuals are evaluated frequently, the same test is used so often that the subject may begin to learn the stimuli. Thus, it is important to have several word lists so that the subjects' scores are not a reflection of their ability to remember the stimuli, but, rather their perception of the stimuli. These alternative forms, however, must have equal difficulty between each list (less than or equal to 8% variability) so that the variations in scores acquired mirror the relative ability of the listener to recognise speech and are not due to difference in difficulty between lists (Markides, 1987).

According to Dillon and Ching (1995), the lists of a speech test are equivalent if any list would result in the same score as any other list when tested under the same test conditions. To achieve list equivalence, test items need to be distributed among the lists such that the items in each list have similar redundancy, phonemic balance and word familiarity. One approach to achieving list equivalence is to use the same words in every list with only the order changed. For meaningful stimuli, this approach only seems suitable when the speech test is to be used to determine the speech recognition threshold (SRT), rather than the maximum attainable intelligibility, because learning of the stimuli by the subject is inevitable.

An example of a test that uses meaningful stimuli with the same words in every list is NU-CHIPS (Elliot & Katz, 1980). The authors developed NU-CHIPS to obtain maximum discrimination (suprathreshold level) with a sensation level of at least 30 dB. Unlike the findings of Dillon and Ching (1995), Elliot and Katz found that learning effect did not occur when the same words in every list with different randomisation were presented to a listener in the same test session.

INDO-SPRITT has four alternative forms and uses the approach where each list has the same words with only the order changed.

2.2.3 Response factor

The type of response needed must reflect the child's ability to hear and not his/her speech production or his/her mental, physical, linguistic or educational abilities. An oral response is inappropriate for those children with severe articulation problems because the tester cannot be sure whether such a response is due to faulty hearing or faulty articulation, or both (Markides, 1987). Similarly, it is inappropriate to accept a written response because this may reflect the child's ability to write correctly rather than his/her ability to hear. The response criteria can be met by using pictures and/or toys to be identified by the child in the identification task. In such a task the child simply points to one of several pictures or toys associated with the word heard, instead of repeating or writing it (Kendall, 1954). Since the subject selects from a small set of allowed responses, these tests are called closed response tests.

Black (1957) stated that multiple choice or closed set response mode produce higher scores than those acquired with open set and/or write down responses, and a reduction in score dispersion. This result can be explained by the effect of a closed set, which can limit the alternative response availability and can be expected to produce higher scores than an open set. For this reason it is crucial to determine the characteristics of the test. If a foil does not represent possible response confusion, the chance probability of target word selection is increased (Mackie & Dermody, 1982).

There are two types of task domain in closed set tests, unrestricted task domain and restricted task domain (Kirk et al., 1997). A task domain is considered *unrestricted* when the test items are not uniquely specified for the listener, but are inserted among foil items representing selected phonemic confusion and an effective foil is a word that sounds similar to the test item (Black, 1968). An example of this type of test is the Northwestern University Children's Perception of Speech Test (Elliot & Katz, 1980).

However, it is interesting to note that the matching words in the NU-CHIPS are based on very inconsistent criteria for the selection of test item and foil (Appendix A). For example, it might be just the vowel that is consistent between the test item and the foils or it might be the last or the first consonant which is consistent between the test item and its foil or it might be more than one sound or phoneme that is consistent between the test item and its foil. Furthermore, some test items have foils that bear no relations to the test item. Thus, the phonetic relationship between the test items and the foils is very inconsistent.

If the criterion for the selection of foils is based on the presence of matching vowels, NU-CHIPS has less test items which have three such foils and more test items that has only one such foil and some of the test items in NU-CHIPS have no such foil. For example, Test Item 23 in List A1 is 'witch' has three foils (fish, sink, watch). Two of these have a similar vowel and one has a similar first consonant. In List A2, 'witch' is Item 29 and the foils are 'fish', 'sink' and 'milk'. Here, the vowels are similar to the vowel in the test item. In List B3 where 'witch' is Item 30, the foils are identical to those in List 2, but in List B4 'witch' is Item 29 and the foils are 'fish', 'frog' and 'watch'. In List B4, one of the foils has a similar vowel, one bears no relation to the test item and one has a similar initial and final consonant. This raises a question whether the variation in the number and type of foils affects the issue of equal difficulty between each list and/or between the test items. Furthermore, if a foil does not represent possible response confusion, the chance probability of test item selection is increased (Mackie & Dermody, 1982; Kirk et al., 1997).

On the other hand, in a restricted task domain, the test item stands alone and the foil items do not represent possible response confusions (Kirk et al., 1997). One example of a test using a restricted task domain is the ANT developed by Erber (1980).

In an unrestricted task domain, the listener must identify the test item in a bottom-up processing strategy or in other words, the listener uses a physical auditory analysis to identify the stimulus, triggered by the arrival of sensory information to the ear. A listener with a hearing problem would have limited information arising from the incoming speech signal, and will not be able to identify the test item on the basis of cognitive abilities. Therefore, it might be said that an unrestricted task domain will increase the sensitivity of the test. In contrast, a test with a restricted task domain may be less sensitive. This is due to the fact that cognitive (top-down) processing is facilitated by a listener's general knowledge (including language and vocabulary level) and by expectation based on the context of the sensory event. Thus, even when only limited acoustic or phonetic information is available for an individual with hearing loss, the listeners may still be able to complete the task using cognitive processing. Therefore, the chance for correct identification is increased (Kirk et al., 1997).

INDO-SPRITT uses picture pointing task in a closed response set. This test format is more suitable for use with very young children rather than repeating the words heard (as in Soewito's procedure). The test items will be inserted among foil items representing selected phonemic confusions (unrestricted task domain) because it is intended to analyse the physical auditory ability of the subject.

2.2.4 Presentation factor

2.2.4.1 Monitored live voice versus recorded

The presentation method of speech discrimination tests for children often uses monitored live voice (MLV) presentation. This method is preferable because it offers the audiologist greater flexibility, which is required when working with children. It allows adjustments to suit the patient's ability and the available test time, either faster or slower presentation, omission of the carrier phrase, or use of selected words. It is also possible that the material is presented to the children in a familiar accent. The audiologist can reinforce when indicated, use encouragement, or change instructions. Basically, it permits the audiologist to do what is necessary to elicit responses. One example of the use of MLV presentation is in the Kendall Toy Test. Live voice testing may also be useful in developing countries where equipment or calibrated equipment may not be available. However, the drawback of MLV presentation is a lack of standardised stimulus materials (e.g. speaker differences).

Recorded speech can also be used for administering speech discrimination tests for children. The benefit of using this format is to ensure that the test material presented is always the same. It permits us to edit a master recording and to approach the ideal of a set test materials of equal difficulty, which give reproducible discrimination curves.

2.2.4.2 Speaker characteristics

Effect of male and female voices

Male and female voices influence the listener's intelligibility because female voices are usually softer and occupy higher frequency ranges than men's voices (Fletcher & Steinberg, 1929) and male voices are about an octave lower in pitch than women's (French & Steinberg, 1947). In the presence of high level noise, the intelligibility of female voices was less than for male voices when the speakers were untrained, however, when they were trained these differences were eliminated (Silverstein et al., 1951).

Further studies (Silverstein et al., 1951 & Palmer, 1955) suggested that the physical differences inherent in male and female voices may have little effect on their intelligibility scores in either group. They used male, female and children's voices to record phonetic balance words, which were amplitude normalised and presented to normal hearing and hearing impaired subjects. Another study found that the average P-I functions for phonetically balanced monosyllable lists using male and a female speaker were the same (Macrae et al., 1963).

Between and within speaker differences

The variability between speakers is contributed to by factors in speech production which include vocal force or intensity, pitch, duration and CV ratio articulation and voice quality (Hirsh et al., 1954; Fry, 1955; Fairbanks & Miron, 1957; Asher, 1958). It has been concluded that the acoustic characteristics of two or more talkers are sufficiently different to cause variability in listener performance when the same world lists are used. (Hirsh et al., 1954; Asher, 1958) One consequence of this effect is that for development of speech intelligibility tests, Carhart (1965) stated that test results using speech material presented by different talkers are not comparable unless their equivalency has been demonstrated. Thus, the use of recorded test materials is recommended since this method can standardise the stimulus material.

One of the factors that cause variability among scores for speech intelligibility within a single speaker is intensity differences. This variability was demonstrated by Brandy (1966) by examining a single speaker in a live and a recorded testing setting. In live testing, a single talker varies in his/her production of the same word on different occasions producing significant variation in listeners. Brandy attributed the variability produced by live voice testing to the changes in intensity of the individual words when spoken on separate occasions. This result therefore supported the use of recorded tests since this method produced more stability. Brandy also stated that a single speaker will not articulate a word in precisely the same way on different occasions. This discrepancy produced differences of up 10% in scores when recorded on different occasions using the same listeners, lists and speaker.

Speaker dialect

Since the measurement of intelligibility involves speech production and speech perception, the influence of dialect and the specification of dialectal differences involve both articulation of the sound and audition. Thus, if a speaker of a particular dialect of English records a list of words, it has been claimed that his/her speech patterns will have a considerable influence on intelligibility. Also the listener’s ability to interpret different dialects and his/her familiarity with the speaker’s dialect will influence the intelligibility of the speaker. As a consequence of this fact, speech test materials should be recorded by a speaker with the same dialect as the listeners or in a general dialect familiar to the listeners from media exposure.

It is intended that INDO-SPRITT will use recorded test material with an Indonesian female speaker who speaks Indonesian language for hers daily communication. However, flexibility in administration will be permitted (e.g. monitored live voice would be used if a child is unwilling to use headphones).

2.2.4.3 Presentation level

Factors that influence the ability to understand speech are not only the listener's hearing sensitivity and the difficulty of the material, but also the intensity of the speech as it reaches the ear (Davis & Hudgins, 1946). The most essential data that indicates an individual's potential recognitions are elicited from a P-I function. These functions are obtained by testing recognition at several different levels (six levels have been reported as optimal) above the speech reception threshold. The level at which the subject performs his/her maximum discrimination is called the PB-Max (Penrod, 1994). A complete P-I function is rarely undertaken because of time constraints. There are several alternative assessment methods that include methods that employ an estimation of PB-Max score which include presentation of the stimuli at a comfortable loudness level or a number of different fixed presentation levels (Mackie & Dermody, 1982). These procedures are non-adaptive or the method of constant stimuli. The majority of speech recognition tests such as AB-Word lists, NU-CHIPS and WIPI are carried out using these methods.

Another method that is often used to administer speech tests is the adaptive procedure which involves changing the presentation level of the stimulus depending on the subject's performance to the preceding stimulus ((Mackie & Dermody, 1986). Adaptive procedures are efficient at targeting the reception or detection threshold of an individual, because trials are placed at a targeted point on a psychometric function, which does not have to be estimated before testing begins. In the case of adaptive speech testing, threshold is defined as the signal level at which the probability of a correct response is halfway between perfect performance (100%) and chance performance (Kaernbach, 2001).

Adaptive procedures have been utilised in clinical audiology since the field began. There are three general categories of adaptive procedures that are commonly employed in current psychophysical testing procedure – staircase procedures, maximum-likelihood procedures and Parameter Estimation by Sequential Testing (PEST). The most common form of the adaptive procedure used in clinical audiology is the staircase procedure (Leek, 2001).

The adaptive staircase procedure was the first adaptive procedure to have a clinical application in audiology, when it was employed by Hughson and Westlake (1944) to determine auditory thresholds for pure-tones. A modified version of this original method (Carhart & Jerger, 1959) is still in used in clinics today. The method involves lowering the stimulus intensity by 10 dB with every detection of a tone, until the subject no longer responds. The stimulus intensity will then be increased 5 dB steps until a response is made.

In the case of speech testing, there are three forms of staircase procedures (Leek, 2001) which include Simple Up-Down Staircase (Dixon & Mood, 1948), Transformed Staircase (Levitt, 1971) and Weighted Up/Down Staircase (Kaernbach, 1991).

INDO-SPRITT will use the simple up-down staircase procedure because it is the most straight-forward method (Mackie & Dermody, 1986) and targets the 50% correct level on a psychometric function (Dixon & Mood, 1948, Levitt, 1978). The simple staircase method involves the reduction of stimulus intensity after every correct response (positive response) and the increase of stimulus intensity after every incorrect response (negative response). The amount by which the level of the stimulus is increased or decreased is referred to as a *step* and the step size is identical for increases and decreases in intensity (Figure 2.1). A series of steps in either the positive or negative direction is referred to as a *run*.

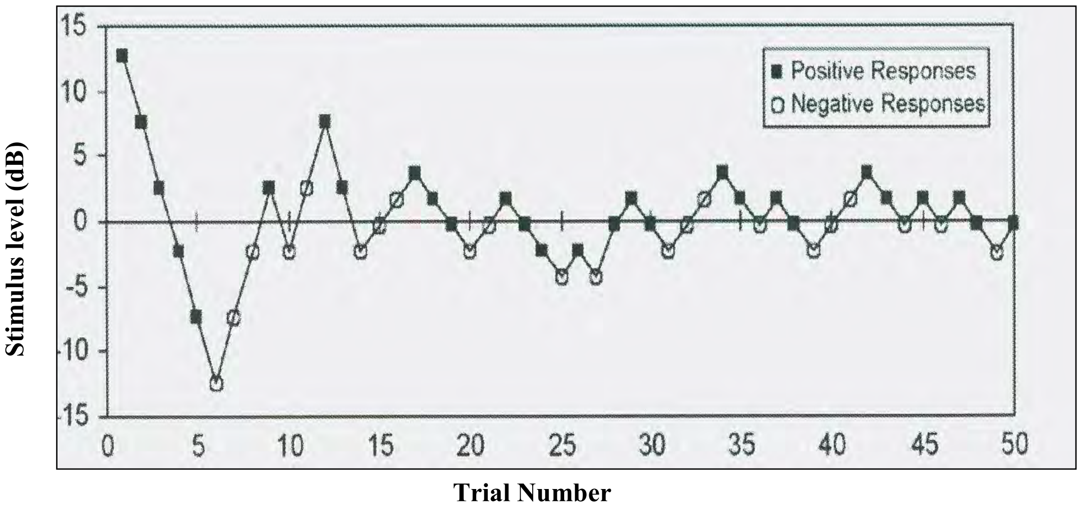


Figure 2.1 Simple up-down staircase (Leek, 2001)

An estimate of the individual's threshold can then be calculated, typically be averaging the levels at the *reversals* (points where the runs change direction). Different researchers used different numbers of reversals to obtain a threshold estimate. For example, Mackie and Dermody (1986) used 15 reversals and made threshold comparison at different points (after 5, 10 and 15 reversals) in the test, whereas, Plant (1990) used 12 reversals.

Furthermore, some studies have used as few as four reversals. However, in many cases, the accuracy and reliability of the threshold estimate may be improved (at the expense of time) by obtaining a larger number of reversals (Bernstein & Gravel, 1990).

To estimate performance at more target levels on the P-I function, Levitt (1971) developed the transformed staircase technique which estimates points higher than 50% correct level. The sequence for a downward movement may be two or more positive responses, and the sequence for an upward movement may remain at one negative response. For example, Levitt says that, two-down, one-up procedure targets the 70.7% level and three-down, one-up procedure targets the 79.4% correct level.

Kaernbach's (1991) weighted up/down staircase can estimate more target levels on a psychometric curve than was possible using Levitt's (1991) transformed staircase procedures. A correct level is analysed according to the desired ratio of up to down steps, and the stimulus level is changed after every trial. Kaernbach described an example of targeting 75% correct performance using a ratio of up to down step sizes of $(1\text{-}p)/p$ or, in this case, .25/.75, or 1/3.

Studies using an adaptive technique in speech perception testing have shown to be reliable when used with normal hearing adults (Levitt & Rabiner, 1967; Bode & Carhart, 1973, 1974), hearing impaired adults (Dirks et al., 1982; Mackie & Dermody, 1982), and with young children (Mackie & Dermody, 1986).

Adaptive procedures offer several advantages over non-adaptive methods or the method of constant stimuli (Mackie & Dermody, 1982, 1986; Lovegrove et al., 1991):

- They eliminate procedural problems found in the constant level approach such as when the test is too easy (scores near 0%) or too difficult (scores near 100%) as a result of the presentation level (they avoid floor and ceiling effects).

- They are more efficient for finding levels on the rising portion of the articulation function and they are flexible in terms of the tester's choice of strategies and target levels.

- They allow standard calculation on an individual on the basis of a single test and therefore produce a quick and reliable estimation of subject variability (i.e it can be expressed as the standard error of the mean).

Many studies, including Ashoor and Prochazka (1985), Mackie and Dermody (1986), Plant (1990) and Nilsson et al. (1994) have used adaptive technique to determine SRT. However, there are other several procedures to obtain SRT, which use ascending or descending methods, such as the ASHA ascending method (1979), Martin and Dowdy's ascending method (1986), ASHA descending method (1988) and Downs and Minard's ascending method (1996), which is a modification of the ASHA ascending method (1979). Some of these methods are similar but take longer than others.

2.2.4.4 Presentation method

It is a common practice that many speech audiometry procedures are presented in quiet, under headphones in an acoustically treated room. However, everyday listening situations are often characterised by noise that interferes with understanding of the speech signal. The type of noise, the position of the signal and the noise, and the signal/noise ratio will vary in each situation for the listener. Thus, the introduction of noise into speech test situation improves the validity of the test by making it a closer approximation to real life (Mackie & Dermody, 1982).

However, INDO-SPRITT was only presented in quiet, because:

- the target population turned out to be severely to profoundly deaf children (Section 2.3) and INDO-SPRITT material was found to be too difficult for this population (explained in Chapters 5 and 6). Presenting the material in noise would have increased the difficulty of the test. As a result of subjects availability, it was decided to develop a test that was suitable for children with profound hearing loss, instead of further developing INDO-SPRITT with subjects with different degrees of hearing loss; and

- there was limited funding and a tight timeframe.

2.2.5 Scoring methods

There are two different methods of scoring: (i) word scoring (synthetic or all or none); and (ii) phoneme scoring (analytic). In the synthetic scoring procedure, the patient must perceive the entire stimulus correctly to receive credit. For example, if the stimulus /cat/ is perceived as /mat/, no credit is given for this response.

On the other hand, the analytic scoring procedure awards credits for any parts of the stimulus perceived correctly. In other words, the scores are obtained by weighing the responses on individual phonemes. Thus, by using the same example, the subject would receive credit for perceiving the phonemes /a/ and /t/ and would be penalised for the incorrect response of the phoneme /m/ (Mendel & Danhauer, 1997).

Thus, phonemic scoring (analytic) is a more informative approach for recording subject responses than the synthetic approach, because phonemic scoring tells the examiner what phonemes are in error (Boothroyd, 1968). Therefore, without increasing the length of the test lists, the phonemic scoring method augments the number of test items and thereby the sensitivity and reliability of the speech discrimination score (Boothroyd & Walker, 1976). However, since the scores in phonemic scoring are obtained by using the responses to individual phonemes, the response method must be an oral or written response. In this case, this type of responses is not suitable for very young children. As mentioned earlier these types of responses might not reflect the child's ability to hear, but the child's other abilities, such as speech production, educational and linguistic ability.

In view of these factors it was decided that INDO-SPRITT' scoring method would be word scoring (synthetic) instead of phoneme scoring (analytic).

In summary it was felt that the speech material used in INDO-SPRITT would need to:

1. be bisyllabic words of BI and representable by pictures;
2. be familiar to the children as young as five years old;
3. be phonemically balanced;
4. have equal difficulty between each list;
5. use a picture pointing task in a closed response set and have foil-minimal pair or contrast in analogous environment;
6. have a recording available with native speaker who used BI for his/her daily communication;
7. permit flexibility in clinical administration (monitored live voice may be used when necessary);
8. be administered to children who use BI for their daily communication;
9. be administered only to children with normal hearing or a mild to moderately severe loss;
10. be administered in quiet, using an adaptive technique; and
11. use a whole word scoring method.

The NU-CHIPS test procedure (Elliot & Katz, 1980) was used as the basis for developing the test for children in BI. NU-CHIPS is suitable for young children and has been found to have a good construct validity and good reliability. As mentioned earlier, it uses a closed response mode using pictures to be identified or pointed at by the child.

This type of response reflects the child's ability to hear and not their speech production or their mental, physical, linguistic or educational abilities.

2.3 Aims of INDO-SPRITT

The original aim in the development of the Indonesian speech test was to provide a means to assess the child's ability to understand speech when presented at a loudness that was well above their threshold (word recognition score (WRS)). The result was to be presented as a percentage score and in view of earlier research there should be a correlation between the type and degree of hearing loss and WRS. Therefore, the speech test was to be used to:

1. assess local children's ability to hear and understand spoken language (maximum percentage score at supra-threshold level);

2. demonstrate to parents an estimation of the child's ability to understand everyday communication- this was felt to be more effectively achieved with a speech test than an audiogram (unaided and aided); and

3. complement pure tone audiometry, which does not assess the effects of linguistic or contextual constraints involved in speech processing.

The original validity assessment was comprised of:

1. familiarity assessment of speech material with normal hearing Indonesian children as young as two years old;

2. reliability measurements including further tests on the equivalence of test forms with different sensation levels and with different age groups, and half-list equivalence;

3. investigation of possible learning effects,

4. performance intensity functions; and

5. studies of hearing-impaired children which would address these questions:

- Are Indonesian hearing-impaired children familiar with the vocabulary of the speech material?
- Do Indonesian hearing-impaired children demonstrate a learning effect when administered more than one test list in succession?
- Are scores for half list presentations equivalent in terms of reliability to full list presentation for Indonesian hearing impaired children?
- How does the performance of Indonesian hearing-impaired children compare with the performance of hearing children?

The target population was children with normal to moderately severe hearing loss.

During the initial field study, it was found that the majority of hearing impaired children found in hearing clinics and hearing- impaired schools in Indonesia were severely to profoundly deaf. This meant that the identification of a population of Indonesian hearing-impaired with a mild to moderate hearing loss was very difficult. They were presumably distributed amongst the population of children with normal hearing in regular schools and would be difficult to include in the study. Thus, given the time frame of PhD study and limited funding, it was very difficult to validate the speech material with varying degrees of hearing loss. Another challenge was the finding of this study, that the children with a severe to profound loss could not achieve any meaningful scores in the familiarity assessment as the test was too difficult for them (Section 4.2.2).

As an alternative, it was decided to use the speech test to provide a means to determine the speech level at which 50% of the presented speech material is correctly recognised by the test subject. This is referred to as the Speech Recognition Threshold (SRT). Like the pure tone threshold-which represents an individual's hearing sensitivity to pure tone signals, the SRT represents individuals' hearing sensitivity to speech signals (Wang, 2006).

The rationale of using the test to measure SRT was because the quantification of hearing sensitivity to a speech signal is often shown by a comparison of the obtained SRT with the normative reference. The normative reference was established with a sample of normally hearing subjects (Doyne, 1951; Soli, 2008). Therefore, validating the test with varying degrees of hearing loss is not necessary.

The validity assessment for INDO-SPRITT involved:

1. the use of speech stimuli for hearing tests which has higher face validity than those using pure tones (face validity);

2. developing lists that are phonetically balanced and of equal difficulty. An assessment was carried out regarding the familiarity of the speech material with normal hearing Indonesian children as young as two years old (content validity); and

3. a comparison between pure tone average (PTA) thresholds and INDO-SPRITT results in order to establish a relationship between the PTA and INDO-SPRITT as a formula for predicting SRT from the average pure tone thresholds (predictive validity).

The reliability of INDO-SPRITT was estimated using three measures:

1. A comparison of the speech threshold at different points of reversals (5, 10 and 15). An increasing difference between thresholds at any of these points reflecting an increase in the subject's variability.

2. A standard error measurement based on midpoints of 5, 10 and 15 reversals.

3. The width of tracking excursion. For each reversal the excursion width was equal to the dB difference between the upper and lower limits of that excursion.

It is intended that INDO-SPRITT can be used:

1. to complement pure tone audiometry;

2. to provide unaided speech threshold and to cross-check pure tone audiometry results;

3. to demonstrate the impact of hearing loss on speech perception to parents and teachers;

4. to provide aided speech thresholds for patients with hearing aids or cochlear implants; and

5. to provide a method of demonstrating aided advantage to parents and teachers.

2.3.1 Complement pure tone audiometry

This proposed test complements pure tone audiometry, which does not assess the effect of higher-level linguistic action or contextual constraint in processing the information, which are involved when listening to speech. The other advantage of undertaking speech discrimination testing with children is that speech items have higher face validity than do non-speech items, since children pay closer attention to verbal than to non-verbal stimuli (Hardy & Bordley, 1951).

2.3.2 Provide unaided speech threshold and to cross-check pure tone results

Soli (2008) suggested the necessary steps to objectively quantify communication for a language are as follows. Once appropriate speech materials for SRT measurements are available, SRTs in quiet and noise are measured with a sample of individuals with normal hearing. This step has been accomplished for a number of languages to date. The mean SRTs in the sample for a particular language provide a normative reference against which the SRTs of hearing-impaired individuals who speak that language can be compared.

Elevated SRTs in quiet and/or noise provide an objective and quantitative measure of communication handicap. SRTs can then, be measured again after medical treatment or hearing aid fitting to determine the extent to which the handicap has been reduced, and to what extent the SRTs have improved.

The key to this approach is the use of adaptive SRTs that are measured and expressed on the same scale for both normally-hearing and hearing-impaired individuals. The performance of normally hearing individuals provides the normative reference used to identify communication handicap (Doyne, 1951; Soli, 2008).

The SRT is also useful as a means to cross-check the validity of pure tone audiometry results, making it a powerful indicator of test consistency (Brandy, 2002; Olsen & Matkin, 1991). The reliability of behavioural data is inferred from the degree of consistency between the SRT value and PTA. As mentioned before, there is generally good correlation between speech and pure tone thresholds both for normal (Fletcher, 1929) and those with hearing loss (Hughson & Thompson, 1942).

The three frequency average of the pure tone air conduction thresholds at 500, 1000 and 2000 Hz which is known as pure tone average (PTA) or speech average (SA) is a good predictor of the SRT in patients with relatively flat hearing loss (Fletcher, 1950; Carhart, 1971). These three frequencies are called 'the speech-range frequencies' because the majority of the speech spectrum occurs within the 500-2000 Hz range. In patients with sharply falling or sharply rising hearing loss, the average of the better two speech frequencies which is known as the Fletcher average (FA) or the two-frequency PTA, or the two-frequency SA is better predictor of the SRT than the three-frequency (Fletcher, 1950).

It is generally agreed that if the relationship between the SRT and the PTA or Fletcher average (FA) are within ±6 dB of each other, there is good agreement; if the scores are between ±12 dB of each other, there is fair agreement; and if the scores differ by ±13 dB or more, there is poor agreement. Disagreement between the SRT and the PTA can be an indication of inconsistencies in the test results. When poor agreement happens and this is not explained by the type of patient (e.g. with a foreign dialect, English as a second language) and/or the SRT (or both) should be retested in an attempt to resolve the discrepancy (Brandy, 2002).

According to American Speech-Language-Hearing Association (ASHA) Committee on Audiology Evaluation (1988), such inconsistencies whereby the SRT is lower than PTA may be due to test variables such as equipment malfunction or misunderstanding of instruction by the participant. Additionally, an SRT that is significantly lower than the PTA can be an indication of pseudohypacusis (Olsen & Matkin, 1991; Lawson & Peterson, 2011), while an SRT significantly higher than the PTA can be an indication of a retrocohlear lesion or some other central auditory disorder (Crandel, 1991), or a language disorder (Silman & Silverman, 1997).

The above values of the relationship between the SRT and the PTA or FA, however, were observed in English language. According to Kruger and Mazor (1987), the relationship between speech threshold and the PTA is dependent on several factors such as the method, configuration of the loss the type of threshold obtained, and also on the speech materials used. Thus, one cannot make the assumption that similar values would automatically be applicable in the Indonesian language without undertaking the relevant research.

In the case of testing a child who does not cooperate with pure tone testing, the SRT can be performed first and this may be the only results obtainable at the first appointment (Lawson & Peterson, 2011). As mentioned earlier, SRT has higher face validity than pure tone and therefore, children might be more interested in doing the SRT test. It may be possible to predict the average pure tone threshold using the established relationship between pure tone and the SRT (INDO-SPRITT).

2.3.3 Demonstrate the impact of hearing loss on speech perception to parents and teachers

Speech audiometry can inform the parents, in a more realistic manner than with pure tones, how an auditory disorder might impact upon the communicative problems in daily life. However, a description of SRT as the speech level at which 50% of the presented speech material is correctly recognised by the test subject may be too difficult for children or parents to understand. A description of SRT as the lowest level at which one can just begin to understand speech may be far more meaningful to patients (Lawson & Peterson, 2011).

2.3.4 Provide aided speech threshold for patients with hearing aids or cochlear implants

Aided speech threshold would be valuable information to see the extent to which the handicap has been reduced, and whether the SRTs have improved (Soli, 2008). However, when establishing realistic expectations to the patients or parents, for the benefits of amplification, one should not be over-reliant on the 50% data point. Instead, obtaining a more complete P-I curve with maximum speech processing performance is a more pragmatic approach for the real-world clinician. A better SRT response is not necessarily associated with better WRS. Using the 50% data point to predict maximum speech processing performance can be misleading, if many other factors are not taken into account (Wang, 2006).

SRT is just the 50% data point on the P-I curve of a subject's speech processing performance. An individual's 50% data point (SRT) on the P-I curve could be the same as another patient's, but the slope and the processing performance of these two patients could be completely different from each other (Figures 2.2 & 2.3).

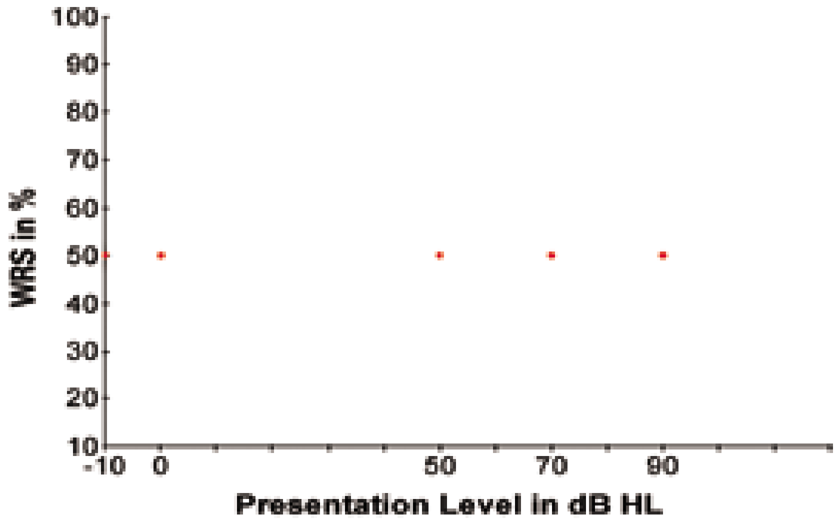


Figure 2.2 50% data point on the P-I curves below (Wang, 2006)

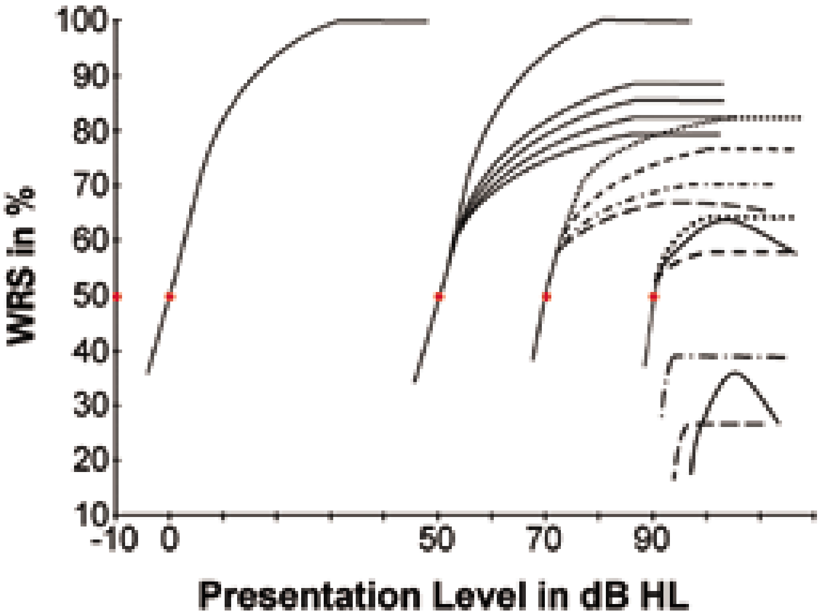


Figure 2.3 Examples of complete P-I curves (Wang, 2006)

The great amount of individual variation revealed from the P-I curves is often related to sensorineural hearing loss (SNHL) as a result of the hair-cell and neural-fibre pathology involved. These losses often feature both a loss of sensitivity and a loss of clarity components.

The clarity loss for speech signals varies significantly depending on the following factors such as:

- degree of hearing loss, shape of hearing loss
- etiology of hearing loss
- pathological condition of the ear-brain structure
- extent of damage to outer hair cells and/or inner hair cells
- damage and effect on active cochlear amplification
- residual function of inner hair cells
- damage to retrocochlear nerve fibres
- effect on neural discharge synchrony
- proportion of retrocochlear lesion versus cochlear lesion
- effect of tonotopic reorganisation of auditory cortex
- length of time the subject has had hearing loss

- history of hearing aid use
- amount of time associated with (in)adequate auditory stimulation
- prelingual versus postlingual cases
- lifestyle and living surroundings
- an individual's linguistic ability

If the pathology involves more damage to inner as opposed to outer hair cells, its effect on signal processing and the amount of distortion during signal processing would likely be greater and higher, because 95% of the auditory nerve fibres carry information from the inner hair cells while only about 5% of auditory nerves innervate the outer hair cells. When the pathology occurs more in the retrocochlear than cochlear region, greater amount of clarity loss and the rollover phenomenon in speech recognition might be expected (Wang, 2006).

2.3.5 Provide a method of demonstrating aided advantage to parents and teachers

It is important for the parents to see the improvement of the SRT of their child after hearing aid fitting or cochlear implantation. However, realistic expectations have to be established about the implication of SRT in quiet in the real world. The SRT indicates the response to speech signals when the speech signal is presented at a fairly soft level such that it is just almost perceptible/recognisable for about 50% of the time. Whereas daily conversations are most of the time performed at a louder level than SRT. The estimation or prediction of how well will the child perform in real life is not comprehensively assessed by SRT testing in quiet.

In summary, with sensorineural hearing loss, the child's aided speech perception ability in the real life may be improved but may not be as good as that of a normal hearing child, especially in noisy environments. However, with conductive hearing loss the result may be similar to SRT with a sensorineural hearing loss, but due to the shape of the P-I function in conductive losses, speech perception at higher levels may be better than that found in sensorineural hearing loss, the child's aided speech perception ability may well be better in real life.

If the child has severe to profound sensorineural loss and/or shows a very poor SRT result (less than 50%), it is recommended that the child should be further tested with INDO-SPASP test (Chapters 5, 6 & 7).

2.4 Research question

The study was intended to answer the following research question:

> **Is the developed INDO-SPRITT test a valid and reliable test that is sensitive to the child test participants' speech perception abilities, at a level which 50% of the presented speech material is correctly recognised by the test subject?**

CHAPTER 3: INDO-SPRITT METHODOLOGY

3.1 Development of INDO-SPRITT materials

3.1.1 Developing word lists

INDO-SPRITT was designed using the basic principles for the design of speech materials, and the special characteristics of Bahasa Indonesia (BI) were taken into account. The word lists in INDO-SPRITT were chosen from twenty nine Indonesian children's books, because a well-established resource of word counts in BI for children was not available. The books consisted of children's story books with pictures for kindergarten and for children at first year primary school. To obtain information about word frequency and phoneme frequency from the children's books, this study used TACT (Text Analysis Computing Tools) software version 2.1.4. This was released in 1995. TACT is owned and managed by the University of Toronto, Canada (http://projects.chass.utoronto.ca/tact/). The program was designed to do text-retrieval and analysis on literary work. Typically, TACT was used to retrieve occurrences of a word, word pattern, or word combination. The program could also be used for simple kinds of analysis, such as sorted frequencies of letters, words or phrases, type/token statistics, or ranking of collocates to a word by their strength of association.

The words from 29 children's books were recorded to calculate the *number of words*. Then the words were assessed by TACT software to obtain *word frequency* data. These frequency values were used to give some indication of the occurrence of words in BI.

The next step was to identify *a corpus of concrete words* (those for which a picture could be used, e.g., *anjing*/dog) which are bisyllabic. As mentioned earlier in the introductory section, the word list used two syllable concrete words because most of the words in BI are bisyllabic.

Fifty of the most frequent concrete words were selected directly from the corpus of concrete words to make up *word lists for the test material*. The selection was based on the following criteria:

- Words with a high frequency of occurrence were used. Ideally the first 50 high frequency words should be used, but in this study that principle could not always be applied, because it was difficult to fulfil all the other criteria (see below).

- To be selected the word had to have at least one similarly sounding word that also occurred at the highest frequency level. If the word in question was not available, a word with the next lowest level frequency was chosen.

- In order to find one or more similarly sounding words. Word structures that were as similar as possible had to be used.

- The 50 words must be phonemically balanced. That means the different phonemes should occur in the test materials in relatively the same frequency as the ones in the sample language from children's books.

An assumption regarding which words were within the receptive vocabularies of three year old children was made at this stage. Because the development of the INDO-SPRITT material was undertaken in Sydney, Australia and there were too few three year old Indonesian children available to check the familiarity of the words. Based on her comprehensive knowledge of the Bahasa Indonesia (of which she is a native speaker) the author made decisions of word familiarity and confirm this judgment based on extensive discussions with Indonesian parents who had 3-year-old children in Sydney. Thus, the words that were judged to be unfamiliar were discarded. Later in the study (Section 4.2.1.1) the list of words was checked with the subjects to ensure that the material were familiar to them.

3.1.1.1 Calculating and balancing the frequency of phonemes of 50 word list in the test material

Counting the frequency of phonemes ensured that the phoneme frequencies used in the test material were similar to the phoneme frequencies occurring in the language. As mentioned earlier, materials are considered as phonemically balanced when the different phonemes occur in the test material at relatively the same frequency as the ones in the language (Egan, 1948). Accordingly, the next step was the counting of *frequency of individual phonemes.*

As mentioned above, the software assessed letter frequency and letter percentage instead of phoneme frequency and phoneme percentage. However, BI has basically a phonemic orthography. Thus, letter frequency data can be used as a basis for analysing phoneme frequency and percentage data, but the letters that were available in the software program did, nevertheless, not represent the entire set of Indonesian phonemes, such as: /ng/, /ny/, /ai/, /au/ and /kh/.

In this case, those phonemes were represented by 3, /ai/ was represented by 4, /au/ was represented by 5, /kh/ was represented by 6 and /ə/ was represented by 7 in the Tact analysis. These frequency values were used to give some indication of the phoneme frequency in BI and the most frequently occurring sounds of BI were used for the test material.

The whole number of phonemes in the existing 50 words was added up in the process, then the percentages of individual phonemes were calculated. After this, the percentages of each phoneme in the word list were compared to the percentages of each phoneme occurring in the language as a whole.

When there was any significant difference between the phoneme frequencies used in the test material and the ones occurring in the language, one or more different words were selected from the corpus of concrete words containing the required phoneme, to match the phoneme frequencies occurring in the language. For example, when the percentage occurrence of the phoneme /ə/ in the test material was only 0.89% and the percentage occurrence of the phoneme /ə/ occurring in the language was 7.41%; in this case more words containing phoneme /ə/ were selected, replacing the words that contained an excessive percentage of phonemes such as the phoneme /u/ (10.71% in the test material and 6.90% in the language). It was also necessary to consider how much discrepancy between the frequency of phonemes in the test material and those occurring in the language were acceptable.

Based on the same 50 words, another three word lists were developed with the same words in a different random order in each list (Appendix B). Thus, INDO-SPRITT has four different test forms.

3.1.2 Developing the foils

The test items are inserted among foil items representing selected phonemic confusions. This is done to let the listener to identify the test item in a bottom-up processing strategy, or in other words, the listeners use a physical auditory analysis, triggered by the arrival of sensory information to the ear. The listener with a hearing problem would have limited information from the incoming speech signal (depending upon the extent of the hearing loss), and such a listener might not be able to identify the test item on the basis of their limited auditory input supplemented by their cognitive abilities. Therefore, unlike a restricted task where the test item stands alone and the foils do not represent possible response confusion, INDO-SPRITT is an unrestricted task domain (because the test items are inserted among foil items representing selected phonemic confusions) and this will increase the sensitivity of this created speech test.

All of the 50 words that had additional similarly sounding words (foils) were selected from the 266 corpus of concrete words.

Foils were chosen based on words with high frequency, phonetic similarity and word structure similarities with the test items and were representable by pictures (concrete word). Foils were chosen based on words with high frequency in the language because if they were not familiar to the children, the foil would not be meaningful and useful.

The problems with the foils in the original NU-CHIPS test have been discussed in Section 2.2.3. In the INDO-SPRITT the foils have been selected with more consistency than in the NU-CHIPS procedure. In all cases, the foils have the same vowel sound as the test item. Also, the foils are the same for every item in each list unlike the situation with NU-CHIPS where they sometimes vary from list to list.

3.1.3 Developing the pictures

The pictures of the 50 words and the pictures of the foils were drawn by an artist based on those used in Indonesian children's books. Two picture response books were developed. Items were arranged so that one book could be used for two test forms (two word lists). Picture Book A is to be used with Word List 1 and Word List 2 while Picture Book B is to be used with Word List 3 and Word List 4. Three pictures representing the foils were selected from the corpus of the 266 two syllable concrete words. At the initial stage of the study, the picture books were in black and white. After familiarity assessment with hearing impaired children completed and further funding was obtained, pictures were redrawn in colour.

3.2 Recording the materials used in the study

Speech materials were recorded in the Macquarie University Linguistics Studio. The speaker was the author (female) who was a native speaker of Indonesia and used the language for her daily communication.

The word lists were recorded with no carrier phrase and were spaced with a six second inter-item intervals. Loudness of each individual word was equalised using the Leq (level equivalent) technique, which means that the words were all at the same root mean square sound pressure level.

The test material words were presented individually to the speaker on a computer screen (using 'QPROMPT' speech elicitation software (Smith, 1987)) and citation forms of each word were collected. When an error or unsatisfactory pronunciation was encountered, the word was presented again until a satisfactory form was collected.

The speech was recorded using an AKG C535EB condenser microphone and placed 150 mm in front and at 45 degrees to the side of the speaker's mouth to avoid turbulent airflow. The frequency response of the microphone is essentially flat between 25 Hz and 20 kHz showing a +/- 1.5 dB ripple above 2000 Hz. The microphone had an equivalent noise threshold of 14 dBA. The microphone was connected to the microphone-balanced, phantom-powered, preamplifier inputs of a Mackie LG 1402 VLZ, produced by LOUD Technologies, Inc (settings: 90 dB S/N ratio flat response). A 64 Hz low cut filter was employed.

Signals were captured through an M-Audio Delta 64 PCI digital recording interface (http://www.m-audio.com) attached to a Microsoft Windows PC. The recording software was 'Cool Edit Pro 2.1' (developed by Syntrillium Software, later acquired by Adobe). Recordings were made at 44.1 kHz sample rate, 16 bit quantisation.

The individual words and announcements were extracted using commercial sound processing software 'GoldWave' (produced by GoldWave Inc., www.goldwave.com).

All speech material was then processed using sound level normalizing and equalizing software 'rmslevel' (developed by R. Mannell, Macquarie University) accurate to within 0.25 dB. The normalizing software does not use frequency filtering. The average level of each speech token is within +/-0.12 dB of a standard reference tone 0 dB @ 1000 Hz.

The audiometer inputs were calibrated with a reference tone 20 dB above the Leq (level equivalent) of the test material. The audiometers used in the study were calibrated for the test input reference level at 0 VU on the meter which, reduces the peak levels to 20 dB above 0 VU, as is normal practice for audiometric speech materials (ISO 8253-3 (2012) Acoustic-Audiometric test methods. Part 3 – Speech Audiometry).

3.3 Field study

The field study involved a familiarity assessment for the material used and a SRT assessment of INDO-SPRITT material to establish the reliability, validity, relationship between pure tone audiometry and the SRT and the normative values of the test.

A Chi-Square test was undertaken to see whether there were significant differences between the occurrences of phonemes in the test material and in the language, both before and after the field study. Chi-square tests enable us to compare observed and expected frequencies objectively. Based on the outcome of the chi-square test we will either reject or accept the null hypothesis (Steinberg, 2011). This study was designed to investigate the relative occurrences of phonemes in the test material and in the language, to be similar. This means that the null hypothesis (that the test materials are representative samples of the language) should be accepted.

A one-way ANOVA was done to firstly analyse the difference between SRT across the number of reversals (5, 10, 15) and secondly to analyse the difference between SRT across age groups. ANOVA was used because it can test the means of more than two groups, and each study had only one independent variable (number of reversals for the first study & age groups for the second one) (Steinberg, 2011. Therefore, a one-way ANOVA was used.

When the ANOVA (*F*) test was significant, a post hoc test was used to locate which groups differ.

A Pearson product-moment correlation coefficient was computed to assess the relationship between SRT and the 3 frequency average hearing level (3FAHL). Pearson (*r*) was used because this is a measure of the linear relationship between two variables that have both been measured on at least on an interval level (Steinberg, 2011).

3.3.1 Familiarity assessment of INDO-SPRITT material

This section will discuss a familiarity assessment with normal hearing subjects and hearing impaired subjects which includes the test items and the foils of the words and the pictures.

3.3.1.1 Familiarity assessment with normal hearing subjects

Word assessment (test item and its foils)

The assessment was initially conducted in a children's hospital. This hospital was a 98 bed private hospital, but all of the subjects were outpatients. However, out of 80 children registered for a hearing test, only 24 participated in this project. Some of the parents reported that they did not feel comfortable with the idea of their child having a hearing test as they had never done it before and others reported that they were reluctant to come to attend the hearing test was due to severe traffic problems around the hospital. In view of this, word assessment was then also conducted in two preschools to increase the subject numbers. Both preschools were private schools. One school had 30 children enrolled and the other one had 20 children.

Subjects

In hospital: 24 children were assessed whose age ranged from two to seven years. They had hearing threshold levels better than 20 dBHL, in at least one ear, at each frequency from 500 to 4000 Hz with tympanometry type A, except two children with bilateral mild to moderate conductive losses.

In two preschools: The ages of 33 children ranged from two to five years and 70% of them were younger than four years old.

A hearing test was not performed in the preschools, because the teachers were not familiar with hearing test and were concerned that the equipment would harm the children. The teachers only allowed the author to perform the test with the INDO-SPRITT picture books and no equipment at all. Even though hearing test was not performed, the teachers reported that all of the children who participated in the study seemed to have normal speech and language development. Thus, although it is realised that this was a far from an ideal situation, an assumption was made that the hearing of these children was normal or close to normal at least in one ear.

The number of subjects within age groups was not even, but depended on the available children at the time of testing and available time of the teachers to allow the children to be tested. Details of the number of subjects within age groups (from the hospital and two preschools) are as follows:

- Two-year-olds: 11 subjects
- Three-year-olds: 25 subjects
- Four-year-olds: 11 subjects
- Five-year-olds: 4 subjects
- Six-year-olds: 2 subjects
- Seven-year-olds: 4 subjects

Equipment

Speech test materials described earlier were used. The audiometric assessments used Interacoustics Audio Traveler AA222, which incorporated both middle ear analyser and diagnostic audiometer and Heine otoscope.

Procedures

The assessment was conducted in a relatively quiet room in all the places mentioned above. A sound level meter was used to ensure that the ambient noise did not exceed 50 dBA. Monitored live voice was not used during the assessment, because the purpose of this study was to assess the familiarity of the word rather than to assess the hearing level of the children.

The child sat about one metre in front of the tester and the words were presented live voice. The child was asked to listen carefully to the word and to look at all the pictures on the page and to point to the picture corresponding to the word that was heard. After the child pointed to a picture, the tester turned the page and asked the child to listen to the next word. The child was encouraged to make a guess even if he/she was not sure of his/her reception and the words (the test item word and its foils) were repeated if the child did not respond within a few seconds.

The words that were highly unfamiliar to many of the 3-year-old and older children, such as *kera* (monkey) and *katak* (frog) were replaced by words with the next lower level frequency of words that sounded similar from the 266 corpus of concrete words. The new words were then re-assessed with the 3-year-old children as subjects. A return visit was made to the schools and the replacement words were tried with 19 of the 3-year-old children who had failed to recognize the words. All of the children were able to recognize the new words.

However, several other words which were unfamiliar to the 3-year-old children could not be replaced, due to the difficulty of finding frequent concrete words in the limited corpus of words in this study. Thus, the inclusion of those words was felt to be necessary and the INDO-SPRITT material is, therefore, more suited for children older than three years.

A statistical analysis was then performed to see whether there were significant differences between the occurrence of phonemes in the test material with the new test items and the occurrence of phonemes in the language (Section 4.2.1.1).

Assessment of pictures (test item and its foils)

When a word was not familiar to a child, it was necessary to investigate the word that was not familiar, or whether it was the picture that the child did not recognise. When a child did not recognise or did not point to a picture or pointed to the wrong picture, for example, *hidung* (nose), she/he was then asked to point to her/his own nose. In the case of words that had a different name for similar/same objects such as *katak* (frog) and *kera* (monkey), many of the children did not recognise or did not point to the correct picture. The children then were asked to point to the picture using the other name such as *kodok* (frog) instead of *katak* and *monyet* (monkey) instead of *kera*. This was done to see whether the problem was the words or the pictures.

The pictures that were not recognised or unfamiliar to the children were then modified.

Later in the process of this study, the pictures (original and the modified ones) were also assessed with 10 adult subjects. This assessment was conducted to see whether the problem that the children encountered was due to the problem of the picture itself or due to context and developmental issues (Section 4.2.1.3).

3.3.1.2 Familiarity assessment with hearing impaired subjects

After testing the INDO-SPRITT material with normal hearing subjects, the modified materials were assessed with hearing impaired subjects. Assessment was initially planned to be conducted with subjects with different degrees of hearing loss. But, as mentioned at the introduction of INDO-SPRITT that the majority of children in many hearing institutions for the hearing impaired had a severely to profound sensorineural loss (Section 4.2.2). Thus, the material was only assessed with subjects who had severe to profound sensorineural hearing loss.

Subjects

Sixteen children who participated in this part of the assessment had hearing aids. Hearing aids were checked prior to the assessment. The age of the participants was between four and 12 years.

The assessment was conducted in a school for the deaf of 80 students in Jogjakarta. All of the students were severely to profoundly deaf. Recent audiograms of the children were obtained from the school principal. At the time of this study, it was not possible for the author to conduct hearing tests as part of the study, because the children had just undergone hearing tests performed by the local hearing clinic. The communication method used in this school was the auditory-oral mode.

Procedures

Initially the procedures were similar to those used with the normal hearing subjects (Section 3.3.1.1). However, none of the subjects could perform the task (familiarity assessment score was close to zero). As a consequence, the subjects were permitted to lip read and most of the time a written version of the word also had to be presented (the familiarity assessment score was better – Section 4.2.2). This was performed because the intention of this part of the study was to assess the children's familiarity of the words and not to assess their sensory ability.

After the completion of testing the familiarity of the INDO-SPRITT materials, modified word lists spoken by the author and recorded by the author. Pictures of the 50 words and pictures of the foils were redrawn in colour by an artist. INDO-SPRITT was then felt to be ready to be used for the SRT assessment.

3.3.2 SRT assessment

INDO-SPRITT was assessed with normally hearing adults and children with normal hearing to mild conductive loss.

3.3.2.1 Assessment with normal hearing adults

It was decided to use adults to determine the efficacy of the test in a more general context because it seemed reasonable to assume that if the test did not work for adults, it would not work for children. It was easier to administer the procedure using more readily available subjects in order to determine the reliability and validity of the test. This should provide an insight into the reliability and validity of INDO-SPRITT. With adults, an open set test was used because adults are able to successfully give a verbal response to an open set design.

Fifteen Indonesian postgraduate university students in Sydney aged between 26 and 51 years volunteered to have their hearing tested. All of the students were native speakers of Indonesian language and used the language for their daily communication. Testing was performed in a sound treated room at Macquarie University at the Paxton Barrand Hearing Clinic and in a quiet room at the Royal Institute for Deaf and Blind Children where the ambient noise level was below 50 dBA.

Equipment

A Madsen OB822 or An Oscilla 960 audiometer, TDH 39 headphones, a GSI 36 tympanometer, a CD player attached to the audiometer and a Radio Shack analogue sound level meter with omnidirectional electret condenser microphone.

Pre test

Prior to the SRT assessment, pure tone audiometry and tympanometry were performed and a list of the words was shown to the participant to ensure that the participant was familiar with the test words. All participants had hearing thresholds better than 20 dBHL and normal tympanograms.

Test administration

The participant was instructed to listen through the earphones to a CD player of a female speaker saying some words. The participant was asked to listen carefully to the word and repeat what was heard. The participant was encouraged to have a guess if unsure.

The adaptive procedure (Mackie & Dermody, 1986) was used. A list of 50 words was tested in each ear using an adaptive procedure. The starting point to assess the SRT was 30-40 dBSL above pure tone average. If the subject gave a correct response, the following stimulus was decreased in 5 dB steps (the first reversal), until the subject gave a correct response. The presentation level was then decreased by 5 dB (the second reversal). At least 15 reversals were performed. The better ear was tested first. A different test form was randomly chosen for each subject and each ear.

3.3.2.2 Assessment with normal hearing children

Fifty-seven Indonesian children (aged 4.5 to 9 years) and four adolescents (aged 10 to 13 years) were involved in this part of the study. All of the children were living in

Jakarta and its surroundings and they all used the Indonesian language for their daily communication.

Testing was performed in a hearing clinic and a kindergarten in Indonesia. The testing room was relatively quiet with ambient noise levels below 55 dBA. The ambient noise was measured using SLM with setting dBA slow. When the ambient noise exceeded 55 dBA, the test was halted until the ambient noise fell below this level. It was initially intended to conduct the SRT assessment in a free field setting at the hearing clinic as a free field presentation may enable the children's parents or teachers to observe at first hand the effects of hearing loss and amplification on the child's speech reception. However, the facilities for free field testing were less than adequate at that time, and as a consequence SRT assessment was only conducted under headphones.

Equipment

TDH-39 headphones with Peltor noise excluding headset, an Oscila 960 audiometer with a CD player attached, GSI 36 Tympanometry and a Radio Shack analogue sound level meter with omnidirectional electret condenser microphone were used.

Pre test

Prior to the SRT test, play audiometry and tympanometry were undertaken with each child. After this, the words with the modified pictures such as *jari* (finger), *kaki* (leg), *hidung* (nose), *guling* (body pillow) and *gayung* (water scoop) were shown to each child to ensure that the modified pictures were easily recognised. All participants had hearing thresholds better than 20 dBHL and normal tympanograms.

Test administration

The child was instructed to listen through the earphones to a CD player of a woman saying some words to the child. Then the child was asked to listen carefully to the word and to look at all the pictures on the page and to point to the picture corresponding to the word that was heard. After the child pointed to a picture, the tester turned the page and asked the child to listen to the next word. The child was instructed to listen carefully since some of the words might be very soft, and the child was encouraged to make a guess if he/she was not sure of his/her answer.

The presentation technique used was the adaptive procedure (Mackie & Dermody, 1986). A list of 50 words was tested in each ear using an adaptive procedure. At the initial stage of the study the starting point of the SRT test was 30 dBSL to ensure that can hear the words clearly. The 30 dBSL was chosen, because Elliot & Katz (1980) found that children of different ages with normal hearing can hear all of the presented words (demonstrated ceiling effects). However, it was observed that 30 dBSL was not loud enough to give reliable answer, compared to 40 dBSL for a few children. Thus, it was decided to start the SRT test at 40 dBSL for the rest of this study.

If PTA results were unknown because the child did not cooperate with the play audiometry test, the starting point was at 50 dBHL, assuming the child's threshold was around 10 dB to ensure the signal was loud and clear. If the child did not respond or gave an incorrect response, the stimulus was increased until the child gave a correct response.

When the child gave a correct response, the following stimulus was decreased by 5 dB (the first reversal), until the child gave a correct response. The presentation level was then decreased by 5 dB (the second reversal). At least 15 reversals were performed. The better ear was tested first, then the test was continued using the other ear. If the child was tired or not able to continue, the test was ended and therefore only one ear was tested. A different test form was randomly chosen for each child and each ear.

Comparison between PTA and SRT was performed and the relationship between PTA and SRT was determined. A comparison of the speech threshold level at different points of reversals (5, 10 and 15) was performed. Standard error measurements based on midpoints of 5, 10 and 15 reversals and width of tracking excursion for each reversal were calculated.

CHAPTER 4:
INDO-SPRITT RESULTS AND DISCUSSIONS

4.1 Development of INDO-SPRITT materials

4.1.1 Word lists

4.1.1.1 Counting the word and phoneme frequency in Bahasa Indonesia

The number of words recorded from 29 Indonesian children books was 6837. Word frequency, phoneme occurrence and phoneme percentage data were obtained from the assessment of results obtained from the TACT software (Bradley et. al., 1995). The word frequencies ranged between 1 and 144, and the phoneme frequencies ranged between 9 (0.03%) and 6745 (19.05%) (Table 4.1). As mentioned above, the software assessed letter frequency and letter percentage instead of phoneme frequency and phoneme percentage. However, BI has basically a phonemic orthography. Thus, letter frequency data was used as a basis for analysing phoneme frequency and percentage data.

The total of 6837 was a relatively very small amount compared to the much larger amount (e.g., 4.5 million words of the Lorg count (Owens, 1961)) that might usually be used in developing speech tests in English. This limitation was caused by the fact that the books used in this study were limited to the ones used in preschools, consisting of words familiar to pre-schoolers. Also, these children's books contained more pictures than words (many words were represented by more than one picture). The small number of books or words used is not the most important issue in the development of this test. The most important issues are that the children can recognise the words, the words can be represented by pictures, and these words will enable us to see how well the children can identify the target speech sounds. .

Table 4.1 Phoneme frequency (by frequency)

Letter	Frequency	%
a	6793	19.05
i	2887	8.10

Letter	Frequency	%
u	2474	6.94
ə	2180	6.11

Letter	Frequency	%
k	2173	6.09
n	2124	5.96
t	1818	5.10
r	1695	4.75
m	1648	4.62
l	1367	3.83
b	1341	3.76
s	1323	3.71
ŋ	1201	3.37
d	1148	3.22
p	1073	3.01
h	939	2.63
o	655	1.84
g	507	1.42

Letter	Frequency	%
e	478	1.34
j	426	1.19
y	320	0.90
c	295	0.83
ny	291	0.82
ai	156	0.44
au	142	0.40
w	137	0.38
f	24	0.07
kh	12	0.03
z	12	0.03
q	10	0.03
v	9	0.03

* Note in the table above: ny, ai, au, and kh each represent 1 phoneme

Total for all letters (Tokens)	=	35663
Total of different letters (Types)	=	36
Type/Token ratio	=	0.0010
Arithmetic Mean	=	990.6389
Standard Deviation (*SD*)	=	1307.0798
Herdan's characteristic	=	0.2199
Repeat rate for all letter 'a'	=	5.25

Two hundred and sixty-six two-syllable concrete words were identified (Appendix C). The frequency of the concrete words ranged between 1 and 49. The concrete words that were used for test material were ones that had frequencies from 4 to 49 (Table 4.2) and each word had at least one similarly sounding word. Words that were unlikely to be familiar to 3-year-old children, even though they had frequencies greater than 4 were not selected. For example, *wajan* (wok) had a frequency of 8 but this word was not used because it was felt that unlikely to be familiar to 3-year-old children. The words which were used in the test materials were checked for familiarity in the field study in Indonesia (Section 4.2.1.1).

This study used words with a frequency as low as 4 because the amount of words were very limited and the criteria such as having similarly sounding words and phonemic balance had to be fulfilled, otherwise there would not have been enough words for the test material. For example, *bebek* (duck) had a frequency of 4 but *bebek* was chosen in order to have enough words with phoneme /e/ and to have a word sounding similar to *ember* (bucket). Words that were unlikely to be familiar to 3-year-old children, even though they had frequencies of greater than 4, were not selected. For example, *wajan* (wok) had frequency of 8 but this word was not used because it was unlikely to be familiar to 3-year-old children. Furthermore, the word *kancil* that had frequency of 14 was not used in this study. This was because in the Indonesian children's books, the word *kancil* was not represented by a picture and therefore no picture of a *kancil* was available for the test material. The word *kambing* (goat) was used instead as a replacement, although the word *kambing* had a lower frequency of 5.

If we follow Owens's study (1961), as mentioned in Section 2.2.1.2, it stipulates the following range of values: Unfamiliar with usage of frequency = 0-100; Moderately familiar = 150-500; Highly familiar = 1000-2000.

One might have the impression that the value of the test material in this study being only 4 to 49 (Table 4.2), belong to the unfamiliar group. However, the conditions under which this study was conducted was different, because this study involved a sample of 6837 words as opposed to the 4.5 million words used in Owens's study. Thus, in this study values between 4 and 49 became representative (Table 4.2).

The reason for having test items that are familiar to the target population is to control the influence of receptive language ability on test performance. It has been shown that items that are not within the patient's vocabulary cause spurious low scores and lead to unnecessary testing, extra expenses and misdiagnosis (McFarlane, 1940).

The most frequently occurring sounds of BI that were included in the test material were /a/, /i/, /u/, /ə/, /k/, /n/, /t/, /r/, /m/, /l/, /b/, /s/, /ŋ/, /d/, /p/, /h/, /o/, /ɡ/, /e/, /j/, /y/, /c/.

Table 4.2 List of 50 words (25 pairs of similarly sounding words

No	Frequency	Words
1	49	Rumah
2	6	Gayung
3	26	Kucing
4	10	Katak
5	11	Hidung
6	27	Tangan
7	11	Badut
8	9	Piring
9	13	Sapi
10	11	Roti
11	26	Bayi
12	13	Jari
13	7	Bantal
14	7	Garpu
15	23	Anak
16	24	Burung
17	11	Gelas
18	10	Nanas
19	25	Pisang
20	5	Kambing
21	16	Bulan
22	6	Ember
23	25	Bola
24	6	Semut

No	Frequency	Word
1	21	Bunga
2	21	Payung
3	9	Guling
4	7	Bapak
5	9	Tikus
6	9	Gajah
7	11	Rambut
8	10	Lilin
9	11	Nasi
10	10	Topi
11	36	Kaki
12	8	Dasi
13	6	Sandal
14	7	Lampu
15	16	Ayam
16	27	Buku
17	33	Kera
18	18	Mata
19	44	Ikan
20	8	Anjing
21	17	Buah
22	4	Bebek
23	10	Coklat
24	4	Telur

No	Frequency	Words
25	5	Gunting

No	Frequency	Word
25	10	Kursi

Another three word lists were developed based on the same 50 words, with the same words in a different random order in each list.

4.1.1.2 Balancing the frequency of phonemes of the 50 word list in the test material

The frequency of individual phonemes ranged from 2 to 44 (Table 4.3). The total number of phonemes in the existing 50 words was 227 and the percentage of individual phonemes ranged from 0.44% to 19.46%. The percentage of each phoneme in the list of 50 words was then compared to the percentage of each phoneme occurring in the language (Table 4.4).

Table 4.3 Phoneme frequencies in the test material

No	Phoneme	Frequency	Frequency of individual phoneme x 100% total frequency of existing phonemes (227) (%)
1	/a/	44	19.38
2	/i/	22	9.69
3	/u/	22	9.69
4	/ə/	4	1.76
5	/k/	14	6.17
6	/n/	11	4.85
7	/t/	10	4.41
8	/r/	11	4.85
9	/m/	8	3.52
10	/l/	9	3.96
11	/b/	14	6.17
12	/s/	10	4.41
13	/ŋ/	13	5.73
14	/d/	4	1.76

No	Phoneme	Frequency	Frequency of individual phoneme x 100% total frequency of existing phonemes (227) (%)
15	/p/	8	3.52
16	/h/	3	1.32
17	/o/	4	1.76
18	/g/	6	2.64
19	/e/	4	1.76
20	/j/	2	0.88
21	/y/	2	0.88
22	/c/	2	0.88
	Total	227	

Table 4.4 Comparison between percentage of phoneme frequency in the test material and the ones occurring in the language

Phoneme	Phoneme frequency in sample of language (%)	Phoneme frequency in test material (%)
/a/	19.05	19.38
/i/	18.10	9.69
/u/	6.94	9.69
/ə/	6.11	1.76
/k/	6.09	6.17
/n/	5.96	4.85
/t/	5.10	4.41
/r/	4.75	4.85
/m/	4.62	3.52
/l/	3.83	3.96
/b/	3.76	6.17
/s/	3.71	4.41
/ŋ/	3.37	5.73
/d/	3.22	1.76
/p/	3.01	3.52
/h/	2.63	1.32
/o/	1.84	1.76

Phoneme	Phoneme frequency in sample of language (%)	Phoneme frequency in test material (%)
/g/	1.42	2.64
/e/	1.34	1.76
/j/	1.19	0.88
/y/	0.90	0.88
/c/	0.83	0.88

A Chi-square test was done to see whether there were significant differences between the occurrence of phonemes in the test material and in the language. The result indicated no significant differences between the two groups (Tables 4.5 & 4.6).

Table 4.5 X^2 Test calculation

	Observed *N*	Expected *N*	Residual
1.00	44	44.0	.0
2.00	22	18.7	3.3
3.00	22	14.1	7.9
4.00	4	16.0	-12.0
5.00	14	13.1	-.1
6.00	11	13.8	-2.8
7.00	10	11.8	-1.8
8.00	11	11.0	.0
9.00	8	10.7	-2.7
10.00	9	8.9	.1
11.00	14	8.7	5.3
12.00	10	8.6	1.4
13.00	13	7.8	5.2
14.00	4	7.4	-3.4
15.00	8	4.0	1.0
16.00	3	6.1	-3.1
17.00	4	4.2	-.2

	Observed *N*	Expected *N*	Residual
18.00	6	3.3	2.7
19.00	4	3.1	.9
20.00	2	2.8	-.8
21.00	2	2.1	-.1
22.00	2	1.9	0.1
Total	227		

Comparison of observed and expected occurrence of phonemes in the 50 word lists. The expected value was calculated using the percentage of the occurrence of each phoneme in the language database.

Table 4.6 Chi-square tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	17.040[a]	21	.709
Likelihood Ratio	17.664	21	.670
Linear-by-Linear Association	.203	1	.653
N of Valid Cases	450		

a. 14 cells (31.8%) have expected count less than 5. The minimum expected count is 1.98.

The critical value of X^2 for *df*= 21 and $p = .05$ is 32.7 and the value of X^2_{obs} obtained from the data is 17.04. Thus, the X^2_{obs} is smaller than the critical value, thus the expected and observed frequencies are not significantly different.

As mentioned in the Introduction of INDO-SPRITT, despite opinions that said the relevance of precise fulfilment of phonemically balanced speech perception test material for predicting communicative difficulties in everyday life due to hearing loss is questionable (Dillon & Ching, 1995; Martin et al., 2000), this study attempted to achieve phonemic balance as it was felt that it was a safer option.

Phonemic balance is achieved by having the phonemes of the test material occur in the same overall proportion as they occur in the language as a whole (Egan, 1948). On the other hand, NU-CHIPS used the Lehiste and Peterson method which produced lists with a phonemic balance similar to a sample of 1263 monosyllabic words (selected from Thorndike and Lorg's word count of 1,000,000 words) rather than using the language as a whole. This discrepancy is due to the difference in the word structure of BI. As mentioned earlier the word structure in BI is not a CVC pattern and most of the words in BI that describe concrete objects are bisyllabic. With bisyllabic words, it becomes too complex to follow Lehiste and Peterson's method and moreover, the database of the words is not large enough. This study, however, has achieved the goal of having each list consisting of the same phonemes with the same frequency of occurrence. This is because all lists consisted of the same 50 words.

The occurrence of the phoneme /ə/ in the 50 word list is still very low compared to its frequency in the 6837 word sample of the language. This is due to the lack of concrete words that contain the phoneme /ə/. Most of the phonemes /ə/ in the language in this study come from the prefixes *ke*, *me*, *ber*, *per* and *ter*. On the other hand, concrete words with phoneme /u/ are plentiful in the 50 word list. It was difficult to reduce the occurrence of concrete words containing /u/ in the test material as it would reduce the occurrence of other phonemes too such as /ə/. Words containing /u/, but not containing /ə/, can be reduced and replaced by other words containing /u/ and /ə/ such as *penyu* and *jeruk*. Unfortunately, these words have a very low frequency of 3, which is likely to be unfamiliar to the child. Moreover phoneme /ny/ as in *penyu* has also a very low frequency of 0.82.

The least frequently occurring sounds of BI such as /ny/, /ai/, /au/, /kh/, /z/, /v/, /q/, /x/, were not included, because the words containing those phonemes are very rare or borrowed from other languages. If they are used in a clinical test it might cause a misinterpretation of the test result. For example, when listeners are totally unable to perceive a particular phoneme which occurs infrequently in normal everyday speech, the handicap he/she experiences is not as severe as it would have been had the phoneme been a more frequent one (Lyregaard, 1987).

Although there are concerns about some discrepancies between phonemes in the language and the ones in the test material (especially with phoneme /ə/ and /u/), as a group these differences are not likely to have been significant.

4.1.2 The foils

As mentioned in the method section, foils were chosen based on words with high frequency, phonetic similarity, word structure similarities with the test item and representable by picture (concrete word). Ideally, each test item has three foils that met all of the above criteria such as *sapi* (cow), *nasi* (rice), *dasi* (tie) and *bayi* (baby). These four words have a high familiarity or frequency, as well as phonetic similarities in terms of vowel sound and similar word structure: CV-CV and representable by pictures. However, some foils did not fulfil all of the above criteria due to the difficulty of finding frequent concrete words in a limited corpus of words in this study. For example: for the test item *bola* (ball), its foils are *domba* (lamb), *tomat* (tomato) and *coklat* (chocolate). In this example, the foils have phonetic similarities in terms of the vowel sound, but the word structures are different from the test item. The word structure of *bola* is CV-CV, *domba* is CVC-CV, *tomat* is CV-CVC, and *coklat* is CV-CCVC. Another example is the test item *garpu* (fork) and its foils are *lampu* (lamp), *badut* (clown) and *rambut* (hair). The word structure of *garpu* is CVC-CV, *lampu* is CVC-CV, *badut* is CV-CVC and *rambut* is CVC-CVC. Note that having foils with similar word structures to the test item, helps to create a better similar sounding word foils to the test item, compared to having the non-similar ones. However, even though the word structures of the foils are different from the test item, these foils are still considered as good foils or effective foils, because the words have the same matching vowels to the test item. As mentioned earlier, an effective foil is a word that sounds similar to the test item (Black, 1968).

In this study, the development of foils was based on words that have the same matching vowels as the test items. This is because, as shown above, the consonant structure in Indonesian bisyllabic words is very complex (with highly variable patterns) whilst the vowel structure is much simpler (only one vowel to each syllable with only two vowels in Indonesian bisyllabic words). Additionally there are very many more possible consonants than vowels in bisyllabic words. So the pattern of consonants in Indonesian words is very complex because of both the number of consonants and the number of places where consonants can occur.

4.1.3 The pictures

Two picture response books (Picture Book A and Picture Book B), which at this stage, were in black and white. Picture book A was to be used with Word List 1 and Word List 2, while Picture Book B was to be used with Word List 3 and Word List 4.

Using test items that are concrete and representable by pictures is an approach to get an appropriate response method for very young children. In picture pointing response, the child simply points to one of several pictures associated with the word heard, instead of repeating or writing it (Kendal, 1954). Thus, having test items that are representable by pictures and picture pointing response method controls the effect of extra-auditory factors such as physical (articulation) problems and educational (writing) ability.

However, despite the attempt to have concrete test items, some of them such as *buah* (fruit), *anak* (child) and *bapak* (father) and some of the foils such as *nenek* (grandma), *ibu* (mother) and *elang* (eagle) could be considered as abstract rather than concrete. The inclusions of these words were necessary, given the difficulty of finding frequent concrete words in a limited corpus of words in this study. Those words, however, could still be represented by pictures and it was a matter of how those pictures were arranged. For example, *buah* (fruit) should not be placed together in one page with pictures such as *pisang* (banana), *tomat* (tomato) or *jeruk* (orange) because they are all fruits. Instead it seemed more appropriate that they should be placed with other pictures of different items such as *rumah* (house), which is a similarly sounding word but which has little in common with the target word apart from its sound. Thus, the subject in this study could identify the target words unambiguously, in the context of word that they responded to.

4.2 Field study

4.2.1 Results of assessing the familiarity & modification of INDO-SPRITT material with normal hearing subjects

4.2.1.1 The word list

Two words that were highly unfamiliar to many 2-year-old up to 4-year-old children were replaced by the next lower level frequency of words that sounded similar from 266 corpus of concrete words. The two unfamiliar words were *kera* (monkey) and *katak* (frog).

Between 45% and 64% of the above subjects did not recognise *kera* (monkey). *Kera* was the target word and it has an alternative word or form which is called *monyet*. When these children were asked to point to *kera*, they did not point to the correct picture or they did not respond. But when they were asked to point to *monyet*, they pointed to the correct picture. However, *kera* could not be replaced with *monyet* because it did not sound similar to the target word and the design of this test was that the target word was inserted among foils. Furthermore, the number of words that sounded similar or had this type of vowel sounds and also representable by a picture was very limited. Therefore, *kera* was replaced by *cecak* (gecko) which was originally the word used as a foil for *kera* (Figure 4.1).

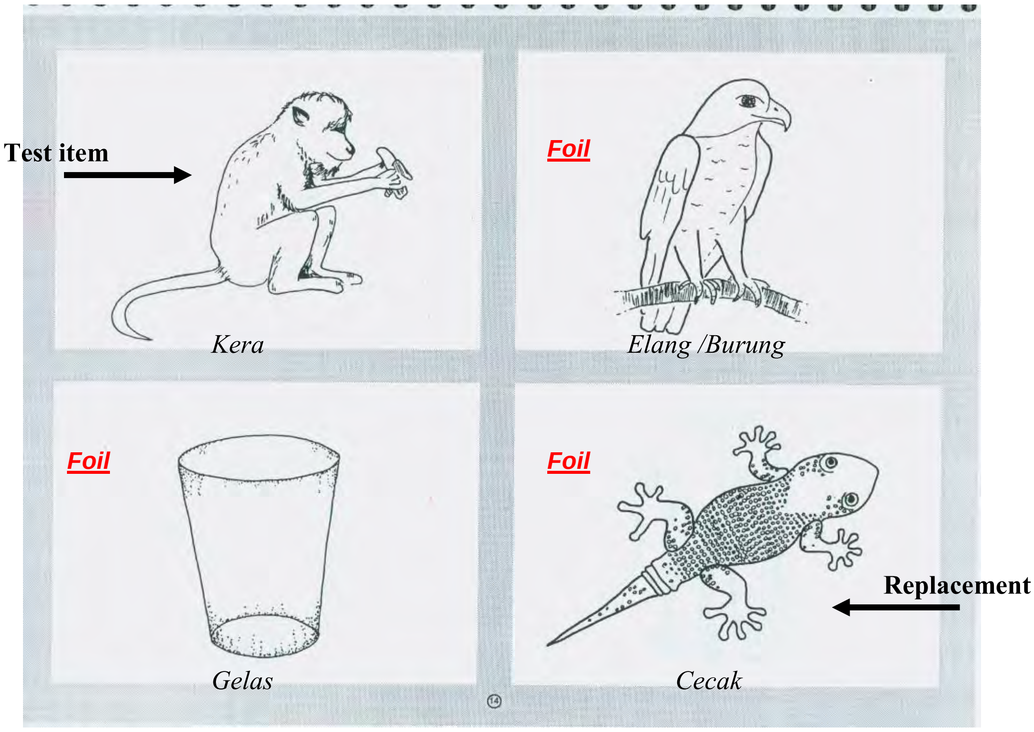


Figure 4.1 Visual representation of test item *kera* (monkey) and replaced by *cecak* (gecko)

Thus, *cecak* became the target word and *kera* became the foil. Unfortunately, *kera* is not a very good foil, because it was not a familiar word and if the children associated the picture with the word *monyet*, the word *monyet* did not sound similar to the foils words. So this test item has less than three foils.

Up to 28% of the subjects did not recognise *katak* (frog) (Figure 4.1a). *Katak* was the target word and there was another name for this particular animal which was *kodok* (frog). When the children were asked to point to *katak*, many of them did not recognise the word and thus could not point to the correct picture. But when they were asked to point to *kodok*, most of them could recognise it. (Only one subject who was four years old could not, recognise *kodok*.)

Katak, however, was not replaced with *kodok*. Because *kodok* did not sound similar to the foils. *Katak* was replaced with *kapal* (boat), which was the next lower level frequency of word that sounded similar from the 266 corpus of concrete words (Figure 4.2b).

Several other words, which were unfamiliar to less than 10% of subjects aged between two and four years were not replaced. The reason was due to the difficulty of finding concrete words of the appropriate frequencies in the limited corpus of words in this study. Thus, the inclusion of those words was felt to be necessary and those words were *bulan* (moon), *gayung* (water scoop), *jari* (finger), *lilin* (candle), *anak* (child), *bantal* (pillow), *nanas* (pineapple), *dasi* (tie), *kaki* (leg) and *kambing* (goat).

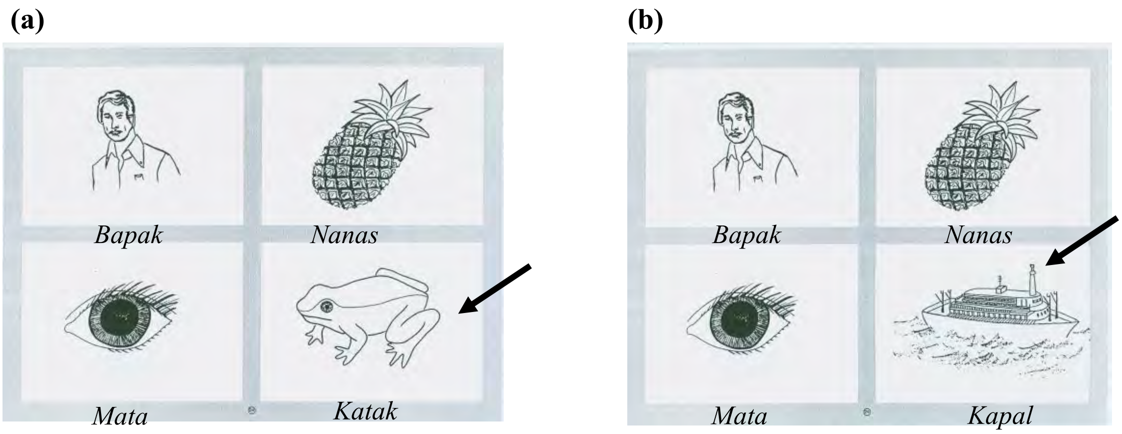


Figure 4.2 Visual representation of test items *katak* (frog) and *kapal* (ship)

Most of the test items in the list of 50 words are still the same as before the field study (Table 4.2) except for *kapal* and *cecak*. (See Table 4.7 for the modified list.)

In a later part of the study, a number of 5-year-olds undertook the test procedure successfully (Section 4.2.3.2).

Table 4.7 List of 50 words (25 pairs of similarly sounding words)

No	Frequency	Words
1	49	Rumah (house)
2	6	Gayung (scoop)
3	26	Kucing (Cat)
4	5	Kapal (ship)
5	11	Hidung (nose)
6	27	Tangan (hand)

No	Frequency	Word
1	21	Bunga (flower)
2	21	Payung (umbrella)
3	9	Guling (body pillow)
4	7	Bapak (father)
5	9	Tikus (mouse)
6	9	Gajah (elephant)

No	Frequency	Words
7	11	Badut (clown)
8	9	Piring (plate)
9	13	Sapi (cow)
10	11	Roti (bread)
11	26	Bayi (baby)
12	13	Jari (finger)
13	7	Bantal (pillow)
14	7	Garpu (fork)
15	23	Anak (child)
16	24	Burung (bird)
17	11	Gelas (glass)
18	10	Nanas (pineapple)
19	25	Pisang (banana)
20	5	Kambing (goat)
21	16	Bulan (moon)
22	6	Ember (bucket)
23	25	Bola (ball)
24	6	Semut (ant)
25	5	Gunting (scissors)

No	Frequency	Word
7	11	Rambut (hair)
8	10	Lilin (candle)
9	11	Nasi (rice)
10	10	Topi (hat)
11	36	Kaki (foot)
12	8	Dasi (tie)
13	6	Sandal (sandals)
14	7	Lampu (lamp)
15	16	Ayam (chicken)
16	27	Buku (book)
17	4	Cecak (gecko)
18	18	Mata (eye)
19	44	Ikan (fish)
20	8	Anjing (dog)
21	17	Buah (fruit)
22	4	Bebek (duck)
23	10	Coklat (chocolate)
24	4	Telur (egg)
25	10	Kursi (chair)

As a result of changing a couple of the test words, the frequency of some phonemes such as /k/, /t/, /r/, /l/, /p/ and /c/ has changed, as shown in Table 4.8 in red colour

Table 4.8 Phoneme frequencies in the test material after the assessments study

No	Phoneme	Frequency	Frequency. of individual phoneme x 100 divided by 228 (%)
1	/a/	44	19.30
2	/i/	22	9.65
3	/u/	22	9.65
4	/ə/	4	1.75
5	/k/	13	5.70
6	/n/	11	4.82
7	/t/	9	3.95
8	/r/	10	4.39
9	/m/	8	3.51
10	/l/	10	4.39
11	/b/	14	6.14
12	/s/	10	4.39
13	/ŋ/	13	5.70
14	/d/	4	1.75
15	/p/	9	3.95
16	/h/	3	1.32
17	/o/	4	1.75
18	/g/	6	2.63
19	/e/	4	1.75
20	/j/	2	0.88
21	/y/	2	0.88
22	/c/	4	1.75
	Total	228	

A Chi-square test was undertaken to see whether there were significant differences between the occurrence of phonemes in the test material after the field assessment (with two new words) and the occurrence of phonemes in the language. The result was no significant differences between the two groups (Tables 4.9, 4.10 & 4.11).

Table 4.9 Comparison between percentage of phoneme frequency in the test material after the assessment study and those occurring in the language

Phoneme	Percentage of phoneme frequency in the language (%)	Percentage of phoneme frequency in the test material (%)
/a/	19.05	19.30
/i/	18.10	9.65
/u/	6.94	9.65
/ə/	6.11	1.75
/k/	6.09	5.70
/n/	5.96	4.82
/t/	5.10	3.95
/r/	4.75	4.39
/m/	4.62	3.51
/l/	3.83	4.39
/b/	3.76	6.14
/s/	3.71	4.39
/ŋ/	3.37	5.70
/d/	3.22	1.75
/p/	3.01	3.95
/h/	2.63	1.32
/o/	1.84	1.75
/g/	1.42	2.63
/e/	1.34	1.75
/j/	1.19	0.88
/y/	0.90	0.88
/c/	0.83	1.75

Table 4.10 X^2 test calculations by category

	Observed N	Expected N	Residual
1.00	44	44.0	.0
2.00	22	18.7	3.3
3.00	22	14.1	7.9
4.00	4	16.0	-12.0
5.00	13	13.1	-.1
6.00	11	13.8	-2.8
7.00	9	11.8	-1.8
8.00	10	11.0	.0
9.00	8	10.7	-2.7
10.00	10	8.9	.1
11.00	14	8.7	5.3
12.00	10	8.6	1.4
13.00	13	7.8	5.2
14.00	4	7.4	-3.4
15.00	9	4.0	1.0
16.00	3	6.1	-3.1
17.00	4	4.2	-.2
18.00	6	3.3	2.7
19.00	4	3.1	.9
20.00	2	2.8	-.8
21.00	2	2.1	-.1
22.00	4	1.9	2.1
Total	228		

Comparison of observed and expected occurrence of phonemes in the 50 word lists. The expected value was calculated using the percentage of the occurrence of each phoneme in the language database.

The critical value of X^2 for $df = 21$ and $p = .05$ is 32.7 and the value of X^2_{obs} obtained from the data is 18.587. Thus, the X^2_{obs} is smaller than the critical value, thus the expected and observed frequencies are not significantly different.

Table 4.11 Chi-square test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.587[a]	21	.612
Likelihood Ratio	19.249	21	.569
Linear-by-Linear Association	.610	1	.435
N of Valid Cases	451		

a. 14 cells (31.8%) have expected count less than 5. The minimum expected count is 1.98.

4.2.1.2 The foils

The ideal number of effective foils is three for each test item. However, the outcome of the familiarity assessment resulted in the situation where some test items had less than three effective foils. This was due to the fact that some foils were not familiar to the children and due to limitations caused by the unavailability of concrete words that met the criteria mentioned in the method section. Therefore, there was a variation of effective foils in INDO-SPRITT material and the variation was as follows:

There are two test items that each has only one effective foil and the test items are *gelas* (glass) and *cecak* (gecko). These test items are presented in Figure 4.3.

As shown in Figure 4.3a, the effective foil for test item *gelas* was only *cecak*. This was because *kera* (monkey) could not be considered as an effective foil, because the majority of the subjects were not familiar with *kera* and it was better known as *monyet* by the children. Furthermore, *monyet* did not sound similar to *gelas*. *Elang* was also not considered as an effective foil because it is better identified as *burung* (bird). Thus, similarly, for test item *cecak* as shown in Figure 4.3b, its only effective foil was *gelas*, because *kera* and *elang* could not be considered as its effective foils.

There are six test items that each has only two effective foils, namely, *topi* (hat) and *roti* (bread); *telur* (egg) and *semut* (ant); *bebek* (duck) and *ember* (bucket).

(a) (b)

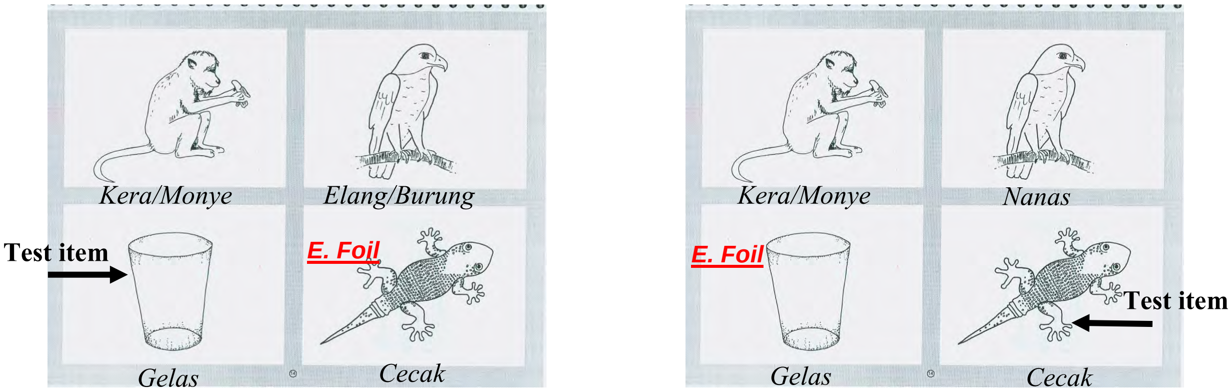


Figure 4.3 Visual representations of test items *gelas* and *cecak* and their foil

As shown in Figure 4.4a, *roti* was the test item, its only effective foil were *mobil* (car) and *topi*. This was because *poci* cannot be considered as effective foil.

(a) (b)

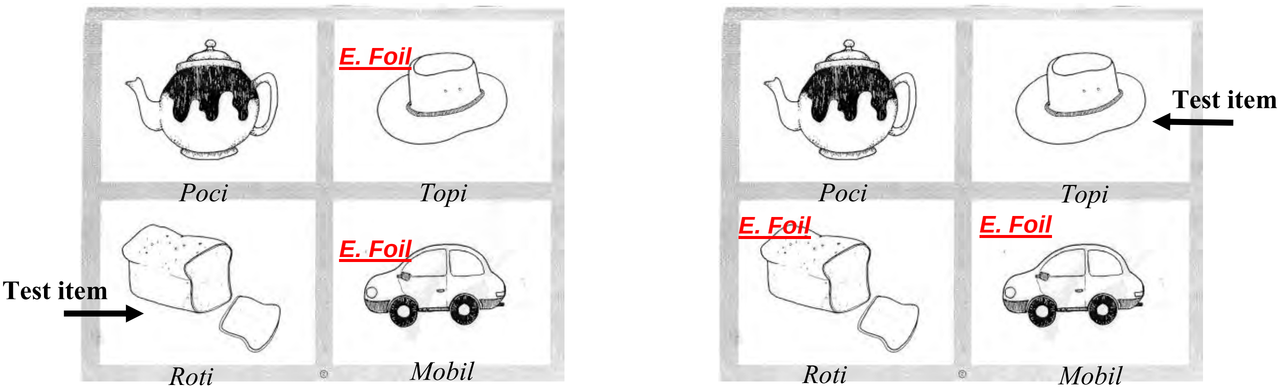


Figure 4.4 Visual representations of test items *topi* and *roti* and their foils

The subjects were not familiar with the word *poci* (teapot) and it was better known as *teko* (teapot). Furthermore, *teko* did not sound similar to *roti* in terms of vowel sound.

Thus, similarly, for test item *topi* as shown in Figure 4.4b, its only effective foil was *mobil* and *roti*.

As shown in Figure 4.5 (a), *telur* was the test item, its only effective foil was *jeruk* (orange) and *semut* (ant). This was because the majority of the subjects were not familiar with the word *penyu* (turtle) and it was better known as *kura-kura* (turtle). Furthermore, *kura-kura* definitely did not sound similar to *telur* in terms of vowels sound and number of syllables. Thus, likewise, for test item *semut* as shown in Figure 4.5b, its only effective foils were *jeruk* and *telur*.

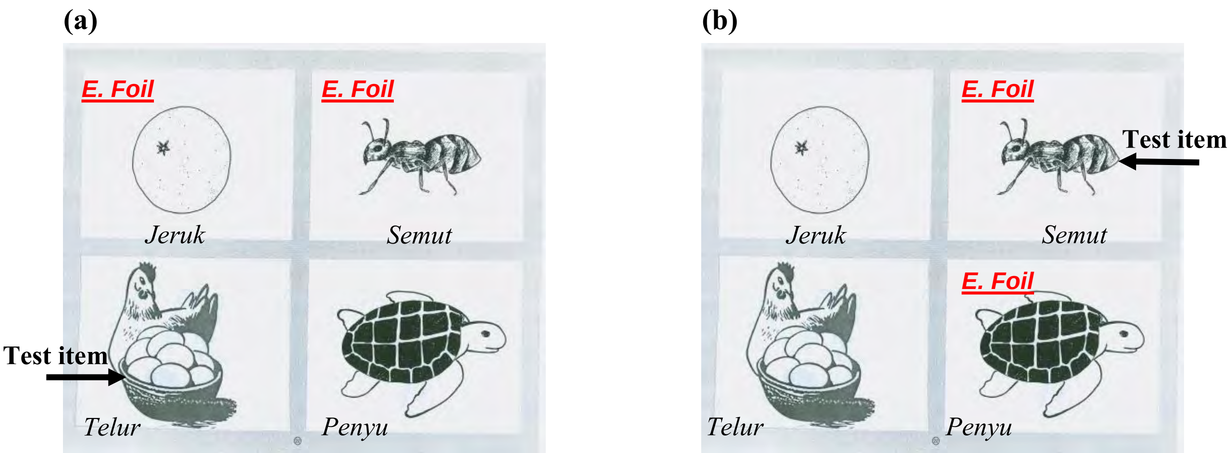


Figure 4.5 Visual representations of test items Semut and Telur and their foils

As shown in Figure 4.6 (a), *ember* was the item and its only effective foils were *nenek* and *bebek*. This was because the majority of the subjects were not familiar with *ceret.* Likewise, when *bebek* was the test item as shown in Figure 4.6b, its only effective foils were *nenek* and *ember*.

Even though, there was variation in the number of effective foils for each test item, the majority of the test items in the INDO-SPRITT have three effective foils and this variation is the same for each list. In summary the variation of foils in the INDO-SPRITT is (i) two test items with one effective foil; (ii) six test items with two effective foils; and (iii) 42 test items with three effective foils (Appendix D).

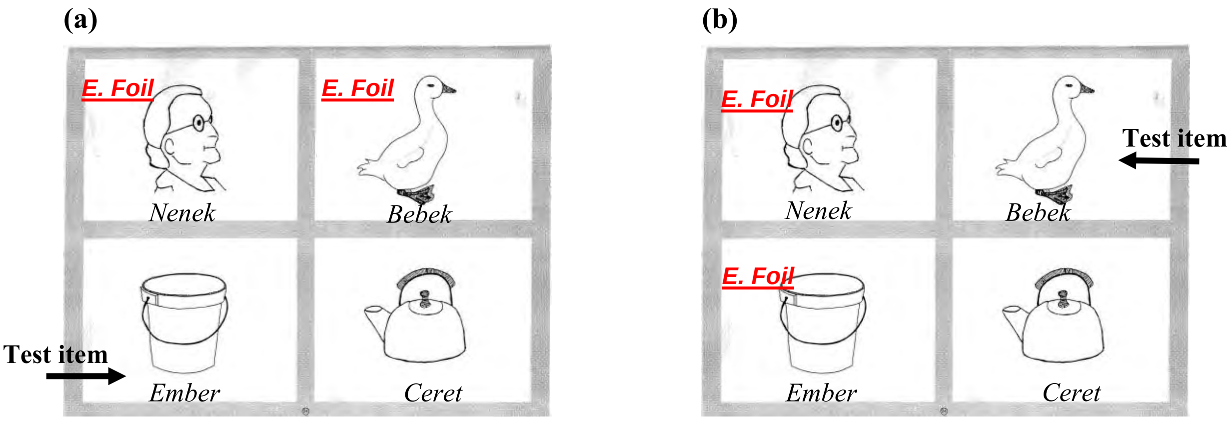


Figure 4.6 Visual representations of test items *ember* and *bebek* and their foils

As mentioned earlier, in NU-CHIPS the selection of test item and foil were based on very inconsistent criteria. Thus, in comparison to NU-CHIPS, INDO-SPRITT materials have more phonologically consistent foils than the NU-CHIPS test material (Section 2.2.3). Furthermore, INDO-SPRITT has clearer equal difficulty between lists than NU-CHIPS, because the variation of foils for a given test item is the same between lists.

The issue of the number of foils is an important one, because it might affect the equal difficulties between test items and between lists. Chance scores will depend upon the number of foils. For example, with one foil the chance score is 50%, for two foils the chance score is 33%, and so on. Therefore, scores may be better or higher for items with fewer foils.

4.2.1.3 The pictures

As mentioned above, during the experimental work when a word was not recognised by a child, sometimes it was the word itself that was not familiar and on other occasions it was the picture that was not recognised by. Several pictures which include *telur* (egg), *hidung* (nose), *jari* (finger/toe), *mata* (eye), *rambut* (hair) and *kambing* (goat) were not recognised or were confused with other alternative pictures by many of the subjects.

Four out of 11 two-year-olds and seven out of 25 three-year-olds pointed the picture of an orange (Figure 5.7a, as indicated by a red arrow) when the word ‘egg’ was heard.

According to parents and teachers, the children are more likely to recognise an egg as a whole egg rather than a broken one even if it is not oval.

The egg picture was then replaced with a whole egg shape in context with the chicken (Figure 5.7b) so it would not be confused with the picture of an orange. All of the children could then point to the correct picture without any hesitancy and none of the children pointed to the picture of an orange. Thus, in this case, the problem was not the word but the picture.

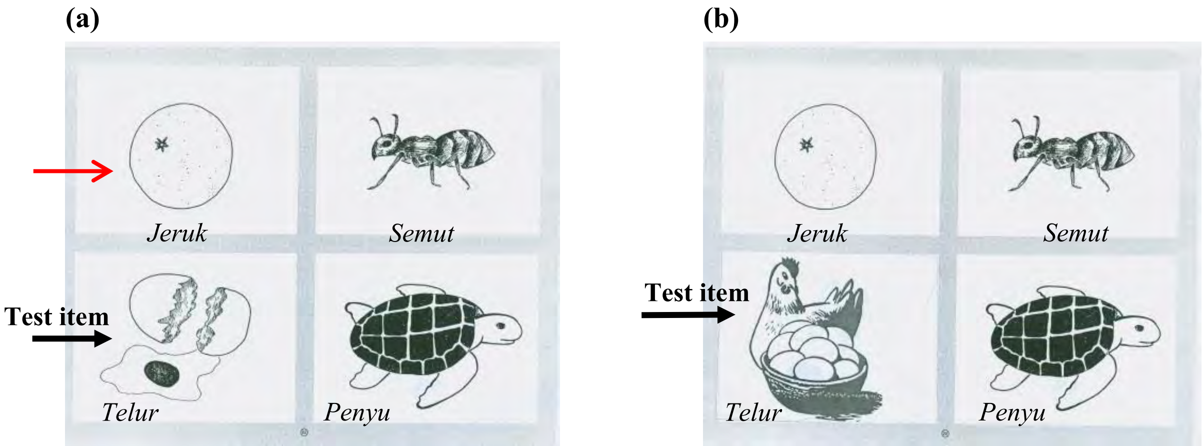


Figure 4.7 Picture of *telur* (egg)

Seven out of 11 two-year-olds, six out of 25 three-year-olds and one out of 11 four-year-olds pointed to *hidung* (nose) in the picture of *ibu* (mother) (Figure 4.8a, as indicated by a red arrow) when the word *hidung* was heard. This indicated that they knew the word *hidung*, but they did not recognise the intended picture of *hidung*. The picture was then modified (Figure 4.8b) and the children could then recognise the picture easily. These children were then asked to point to their own nose and they all could do it. This indicated that the intended picture of the nose was the problem not the word nose.

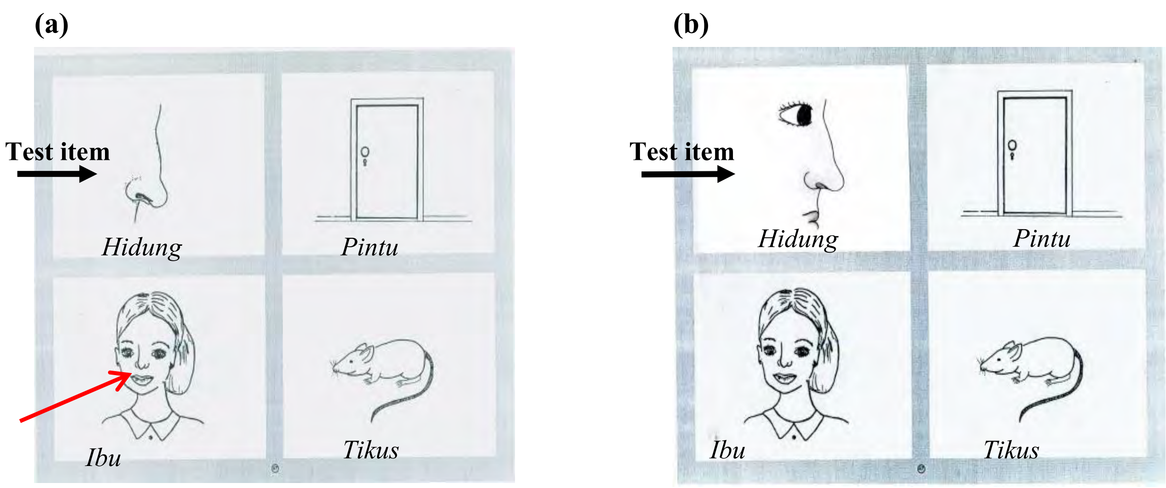


Figure 4.8 Picture of *hidung* (nose)

Two out of 25 3-year-old children did not point to any picture in the book when they were asked to point to the picture of a *jari* (finger or toe). However, when they were asked to point to their own finger or toe they could do it. One out of 11 2-year-olds and three out of 25 3-year-olds pointed to the tip of the foot of the animals or of the human leg (Figure 4.9a and Figure 4.9b, as indicated by the red arrow). We might think that those children were wrong because they pointed to the wrong picture, but, they might be right, because *jari* in the Indonesian language can mean finger or toe. Thus, these children could have been pointing to toes. This was clarified by asking the children to point to their own finger or toe and they could all do it.

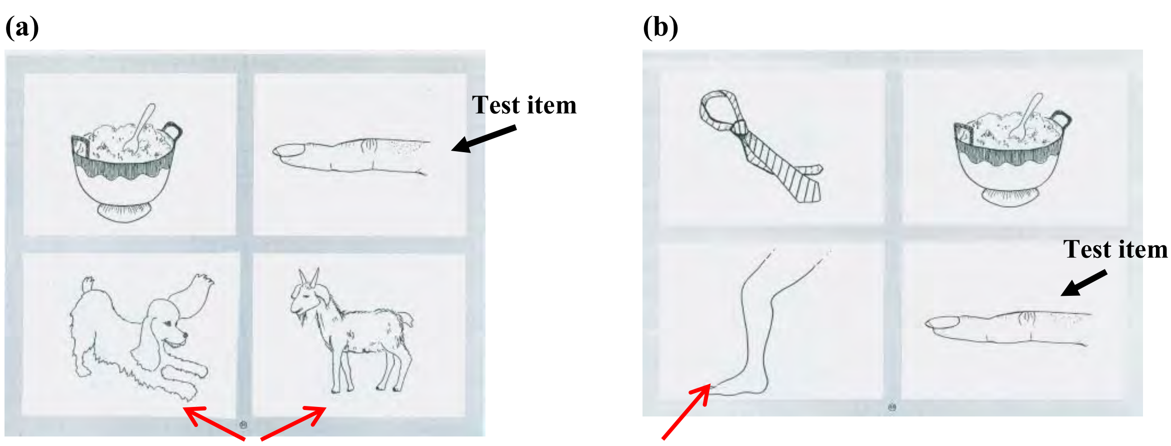


Figure 4.9 Picture of *jari* (finger or toe)

The participants were then asked to point to a picture of *jari* in the book and they pointed again to the same picture (i.e. they pointed to the tip of the foot). Therefore, this picture was then modified, because not only it could be confusing to the children but also to the tester. The modified pictures in Figures 4.10a and 4.10b were then retested and the children could recognise the picture easily and they did not point to the tip of the foot of the animals or of the human leg any longer.

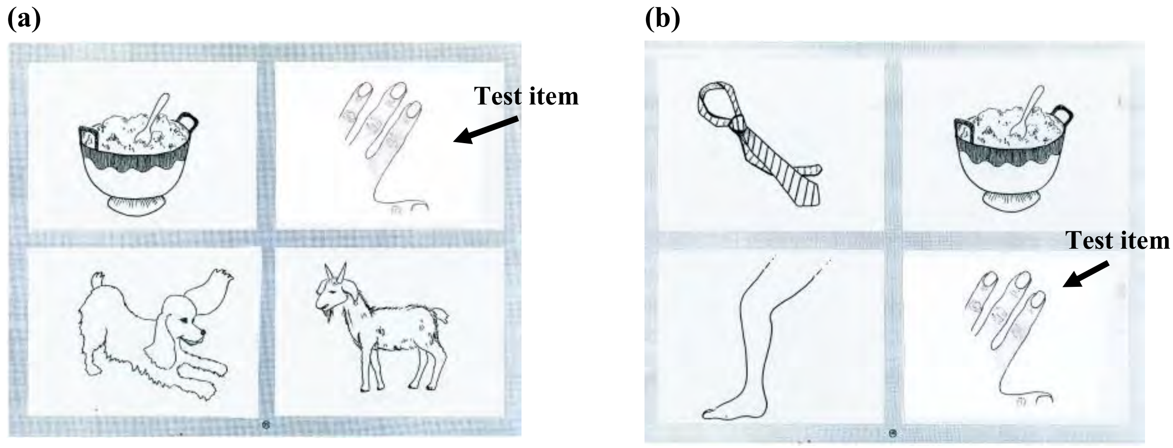


Figure 4.10 Modified picture of *jari* (finger or toe)

Two 2-year-old subjects pointed eye to *katak's* eye (frog's eye) instead of the single eye picture as in Figure 4.11a (red arrow). The rest of the 2-year-old and older children pointed to the intended eye picture. Although only two of the 2-year-old children who were confused by alternative picture, the picture of *katak* was replaced with *kapal* (Figure 4.11b). The replacement by *kapal* was related during the familiarity assessment of the test items. As mentioned before, *katak* (frog) was one of the test items and there is another name for this particular animal which is *kodok* (frog). When the children were asked to point to *katak*, many 3-year-olds did not recognise the word and thus could not point to the correct picture, but when they were asked to point to *kodok*, most of them could recognise the picture.

Katak, however, was not replaced with *kodok* because the design of this test is such that the test item is inserted among foils (words that sound similar to the test item) and *kodok* does not sound similar to the other words/foils. *Katak* was replaced with *kapal* (ship) instead of *kodok* since *kapal* had also been tested with children as young as two years old and it was recognised by them.

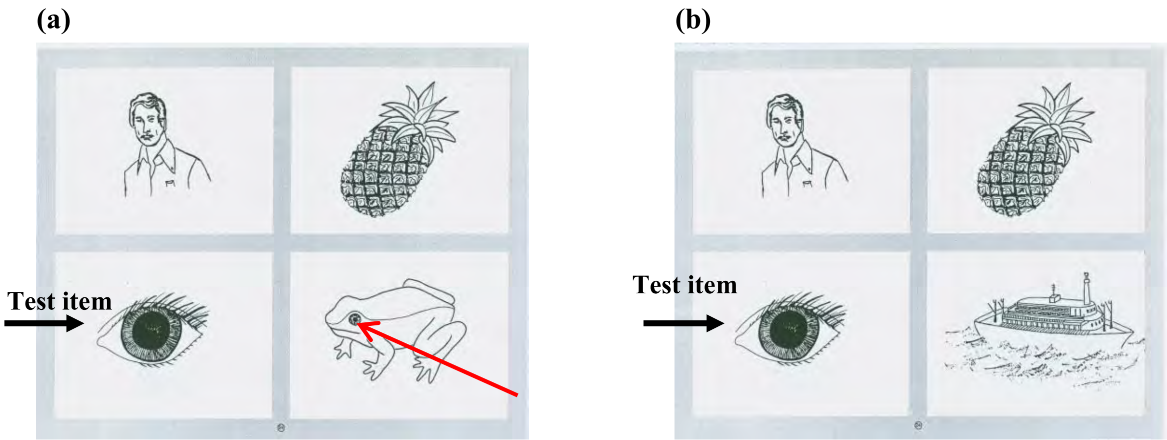


Figure 4.11 Picture of *mata* (eye)

One out of 11 of the 4-year-old children pointed to the *badut* (clown) picture instead of the intended picture of *rambut* (hair) (Figure 4.12a, as indicated by the red arrow). The child was then asked to point to his own hair and he could point to his hair correctly. Then, the child was asked to point back to picture of *rambut* in the book and he pointed again to the picture of the clown. Again, in this case, the tester did not think that this child was making a mistake, because, it was clear that the child was familiar with the word *rambut*.

Furthermore, it was easy to understand why the child pointed to the clown's picture, because the clown also had hair. Although, there was only one child who was confused with this particular picture, this was taken into account for modifying the pictures. Badut's picture was redrawn with a hat and without hair (Figure 4.12b, as indicated by the red arrow) to avoid this confusion. This phenomenon could be due to an issue of clarity design (visual complexity). There were two versions of hair picture, the first one was the hair of the woman and the second one was the clown's hair (Figure 4.12a) and this could cause confusion for the children.

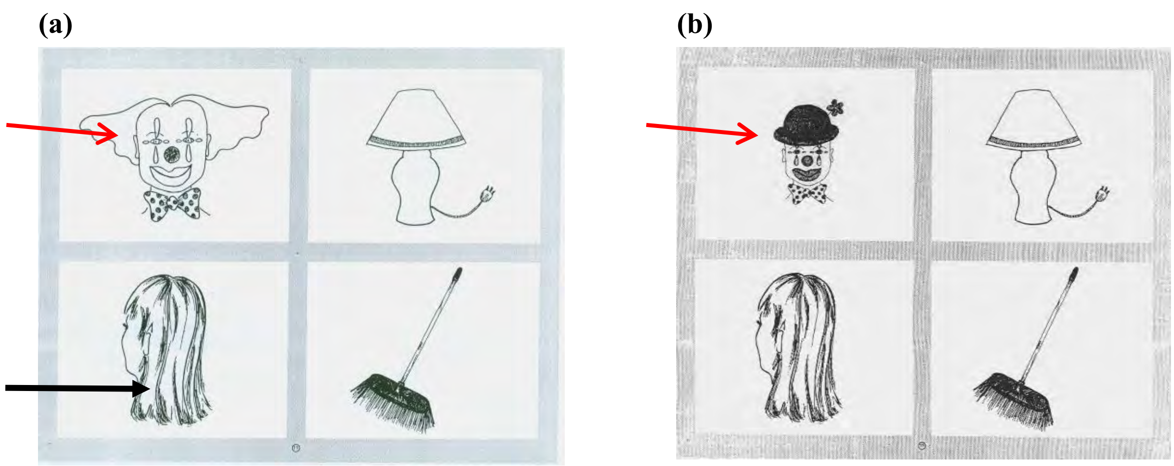


Figure 4.12 Picture of *rambut* (hair)

Four children (one at 2 years old, one at 3 years old and two at 4 year old) got confused with the picture of the test item *kambing* (goat) (Figure 4.13a). It took them longer to choose the picture, they initially pointed to the *anjing* (dog) picture (Figure 4.13a, as indicated by the red arrow) and the *kambing* picture alternately. They were then encouraged to only point to one picture. Three of them chose to point to the *anjing* picture and only one 4-year-old chose to point to the *kambing* picture. It is possible that this phenomenon occurred because the *anjing* picture rather looks similar with *kambing* picture or *anak kambing* (baby goat) (Figure 4.13a).

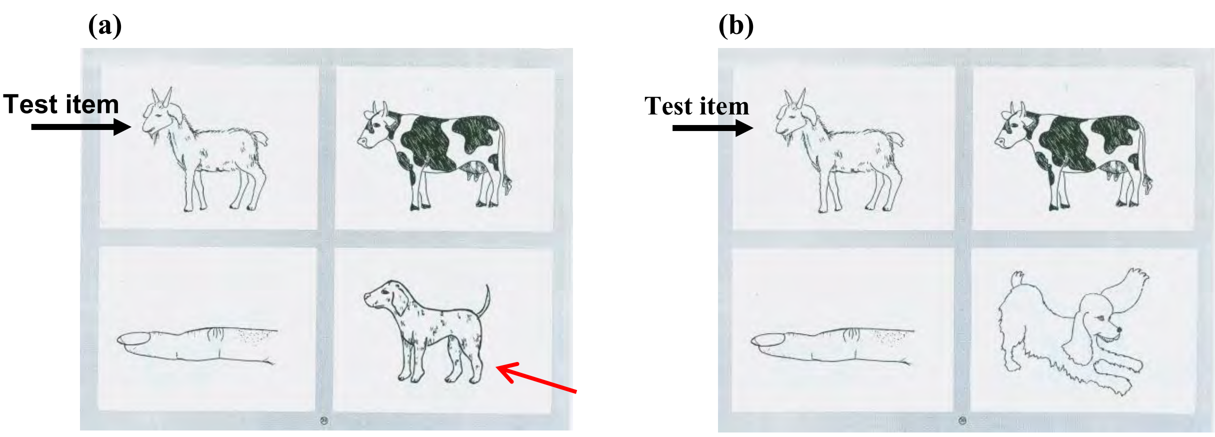


Figure 4.13 Picture of *kambing* (goat)

On the other hand, when the children were asked to point to the *anjing*, all subjects pointed to the correct picture, except one 2-year-old who did not point to any picture.

The *anjing* picture was then modified (Figure 4.13b) and no children were confused and could recognise the picture more easily.

In summary:

- 2-year-olds had problems with seven pictures: telur (egg), mata (eye), hidung (nose), jari (finger), sandal (sandal), guling (body pillow) and kambing (goat)
- 3-year-olds had problems with five pictures: telur, jari, hidung, anjing and kambing
- 4-year-olds had problems with two pictures: hidung and rambut (hair)
- 5-year-olds had problems with one picture: telur
- 6-year-olds had no problems
- 7-year-olds had no problems

The result of picture assessments using 10 adult subjects was that all of the adults agreed with the name of the pictures and they were able to recognise all of the pictures. This indicated that the problems that occurred were because of some of the pictures were not age-appropriate.

Table 4.12 Age effect on picture identification

Age	n	n_p	**%**	
2	11	11	100	most
3	25	13	52	half
4	11	3	27	quarter
5	4	1	25	quarter
6	2	0	0	none
7	4	0	0	none
Adult	10	0	0	none

n = number of subjects; n_p = number of subjects who had problems with pictures

Table 4.12 shows that under the age of four, 50% or more of the children demonstrated difficulty in identifying a picture out of context or an abstract representation of the object, whereas from six years onwards, these difficulties were no longer present.

These phenomena can be explained in three aspects (Wechsler, 2002):

1. Firstly, there is a factor of design clarity with the outline of the picture (visual complexity) and with the details or background. An example is the difficulty identifying or distinguishing the picture of *anjing* (dog) and *kambing* (goat), as mentioned earlier. The design clarity (visual complexity) of the picture was not age appropriate and also due to the lack of life experience of the children who found it difficult to perceive the difference in the tail of the pictures of *anjing* (dog) and *kambing* (goat) and that the goat had horns. But, the 6-year-old children and older had no problem in the task.

2. The second factor is context. The younger the child the more likely it is that they would have difficulty identifying a picture out of context or an abstract representation of the object. It is not just how the representation of vocabulary (drawing) is presented, but also the context in which the vocabulary has been learned, are both of primary importance the younger the child (an example is the difficulty with the picture of an egg mentioned earlier).

3. The third factor is concept learning. Children from 6 years old and older have the life experience to be able to move a concept across contexts, for example, recognizing an eye on the face in a drawing or as a single element. Also they have the cognitive skills to apply information in many different ways. Whereas in children under the age of four, 50%, or more of them, demonstrated difficulties to move a concept across context. For example, these children recognised an eye or nose on a face in a drawing but not as a single element. They lack the cognitive skills to apply information in many different ways.

All these factors will impact upon the ability of the child to recognise a picture and to match a word to it.

As mentioned before in the methodology, at the initial stage of the study the picture books were in black and white in order to be consistent with the picture books in the NU-CHIPS. As the development progressed and further funding was obtained, the pictures were redrawn in colour (Appendix E). This was because pictures in black and white are less realistic than pictures in colour. It seems that younger children need more realistic pictures than older children (Wechsler, 2002).

4.2.2 Results of assessing the familiarity with hearing impaired subjects

As mentioned earlier, the assessment was initially planned to be conducted with subjects with different degrees of hearing loss. But, interestingly the majority of hearing impaired children in many hearing institutions such as the school for the deaf (Karya Mulia) in Surabaya (East Java) (Henderson, 2006), Yayasan Aurica (early intervention support program for the profoundly deaf children and their parents) (East Java), Soetomo Public Hospital (Surabaya, East Java), the two major private hearing clinics (ABDI & Kasoem) that had several clinics throughout Indonesia, and the school for the deaf (Karnnamanohara) in Yogyakarta (Central Java) were severely to profoundly deaf.

Because of issues outside of the control of the researcher, assessment was only carried out at one school (Karnnamanohara).

None of the subjects from the Karnnamanohara school had worn hearing aids consistently because they could not afford to purchase batteries regularly

The modified material of INDO-SPRITT was familiar to hearing impaired children as young as 7 years and 6 months old (Table 4.13). It was felt that a 98% score was an appropriate criterion. There was no confusion with the modified pictures. These children could recognise the modified picture easily once they knew the word. This contrasted with the children who had normal hearing who were able to cope with the test material at a younger age (Section 4.2.1.1)

Table 4.13 Results on the familiarity assessment of the modified material with hearing impaired children

Subjects	Age (years, months)	Score (%) on familiarity of the modified material
S1	4,11	60
S2	5,8	76
S3	5,10	88
S4	6,5	90
S5	7,6	98
S6	7,6	98
S7	8,3	100
S8	8,4	100
S9	8,10	100
S10	9	100
S11	9	100
S12	9,4	100
S13	10	100
S14	10,2	98
S15	10,7	98
S16	12,2	100

As mentioned earlier, a wide range of cues were required for familiarity assessment of the INDO-SPRITT test items with the hearing impaired children. Words were presented orally; the child was permitted to lip read and most of the time a written version of the word had to be presented. Otherwise none of the words could be recognised by the subjects. This could be due to the fact that even though INDO-SPRITT is a closed set speech recognition test, it has similarly sounding concrete words (e.g. *kucing* /cat, *kursi*/chair, *kunci*/key and *gunting*/scissors) that require a picture-pointing response. In this test the child has to select the target word *kucing* among three other similarly sounding words. These words are all bisyllabic, have similar vowel sounds and most of the words and their foils had a similar waveform envelope. Hence, the experiments indicated that, profoundly deaf children who can only hear the vowel sounds or who can only perceive the suprasegmental aspects would produce a poor score as they could not tell the difference between those words.

In view of the result of the experiments, it was concluded that INDO-SPRITT procedure of testing should not be used with severely to profound hearing impaired subjects. INDO-SPRITT procedure is probably best suited for normal to moderately severe hearing impaired subjects. A different assessment procedure was therefore developed for use with severely-profound hearing impaired subjects (Chapter 5). The coloured pictures and the modified word list were retested with Indonesian normal hearing children as young as 4.6 years old (Section 4.2.3.2).

4.2.3 Results of assessing the SRT

4.2.3.1 Results of assessing the SRT with normal hearing adults

Reliability

Comparison of speech thresholds at different points of reversals (5, 10 and 15)

Table 4.14 Mean of SRT at different values of reversals for adults

	5 Reversals	**10 Reversals**	**15 Reversals**
Mean of SRT (dBHL)	17.29	17.69	17.68
Standard Deviation	7.58	7.64	7.66
N	28	28	28
Standard Error	1.43	1.44	1.45
Confidence Interval	2.81	2.83	2.84
Range	14.48 – 20.10	14.86 – 20.52	14.84 – 20.52

A one-way ANOVA was conducted to compare the SRT across the number of reversals (5, 10 and 15) and the result showed that the difference between speech thresholds were not significant at the $p < .05$ level [$F(2,81) = 1.48$, $p = .975$] (Figure 4.14).

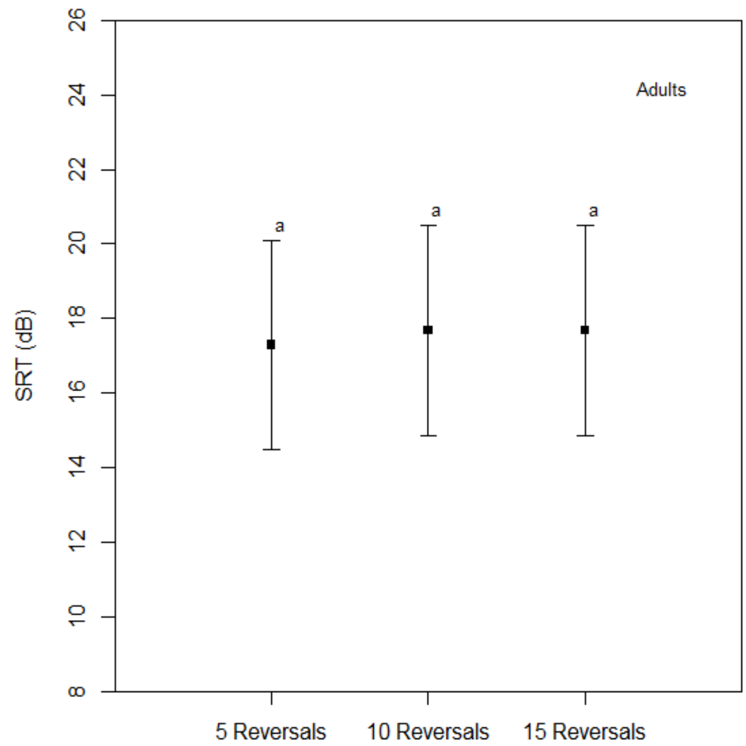


Figure 4.14 SRT for adults using open set speech material (When all groups have the same label "a", this indicates that the three reversals are not significantly different)

Thus, this indicates that 5 reversals is enough to estimate SRT with INDO-SPRITT material without sacrificing the accuracy and reliability of the thresholds. In some studies such as Doyne's (1951) and Soli's (2008), the mean SRTs provide a normative reference against which the SRTs of hearing-impaired individuals can be compared. However, given the number of samples is only 10 adults, the results of the mean SRT in this study is less than adequate to provide a normative reference.

Standard error (SE) measurement based on midpoints of 5, 10 and 15 Reversals

Standard error values were calculated for each subject's threshold after 5, 10 and 15 reversals and the mean standard error for the group at these reversal points is presented in Table 4.15. Results of *SE* measurements (Table 4.15) are all less than 2.5 dB for all reversals and age groups. *SE* of less than half the step size is acceptable (Mackie & Dermody, 1986). The step size in this study was 5 dB. The *SE* decreased over 15 reversals. This trend was to be expected, as an increase in the number of reversals should produce a decrease in the *SE*.

Table 4.15 Standard error measurements for adults

	Number of Reversals		
	5	**10**	**15**
Mean of SRT (dB)	17.29	17.69	17.68
Mean of *SE* (dB)	0.95	0.76	0.70
Range of *SE* (dB)	0.76-1.14	0.66-0.87	0.61-0.79

Width of the tracking excursion

Finally, the width of the tracking excursion was considered as an index for test reliability. For each reversal, the excursion width was equal to the dB difference between the upper and lower limits of that excursion.

Table 4.16 presents the average excursion width. Given the step size of 5 dB used in this study, the results indicate that on average each subject required one word above or below the SRT point to establish a reversal (Figure 4.15).

Table 4.16 Width of tracking excursion over 15 reversals for adults

The width of the tracking excursions	Adults
Mean (dB)	7.54
Standard Deviation (dB)	4.90
N	450
Standard Error (dB)	0.23
Confidence Interval (dB)	0.45
Range (dB)	7.08-7.99

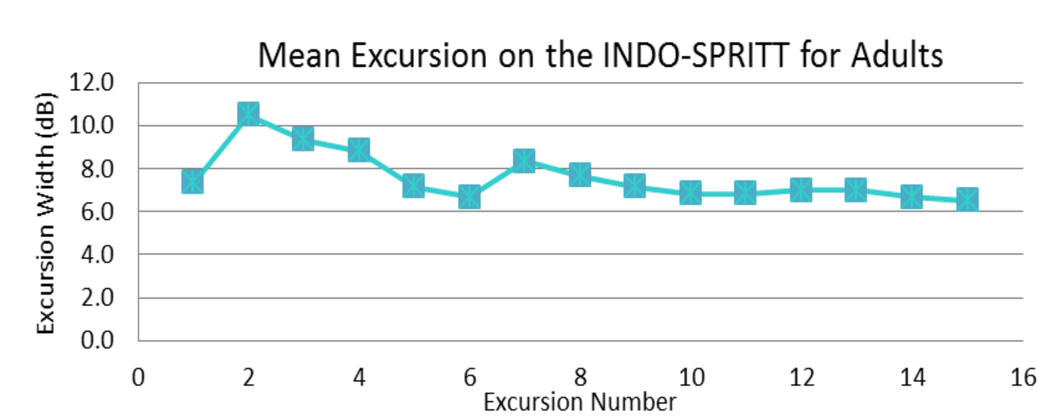


Figure 4.15 Mean excursion on the INDO-SPRITT for adults

Predictive validity

A Pearson product-moment correlation coefficient was computed to assess the relationship between SRT and the 3-frequency average of hearing level (3 FAHL) of 500 Hz, 1000 Hz and 2000Hz (Table 4.17). The results showed that there was no correlation between 3 FAHL and SRT at any number of reversals (Table 4.18). Figure 4.16 shows a non-significant ($p > 0.05$) relationship between 3 FAHL and SRT for 15 reversals.

Table 4.17 Mean SRT and 3 FAHL of adults

Adults	Reversals			3 FAHL
	5	10	15	
Mean (dB)	17.29	17.69	17.68	3.79
STDEV (dB)	7.58	7.64	7.66	4.52
N	28.00	28.00	28.00	28.00
Range (dB)	14.48 – 20.10	14.86 – 20.52	14.84 – 20.52	2.12 – 5.46

Table 4.18 Correlations between SRT and 3 FAHL at 5, 10 and 15 reversals

Reversals	3 Freq avg
5	$r = 0.175$, $p = 0.371$
10	$r = 0.187$, $p = 0.341$
15	$r = 0.213$, $p = 0.277$

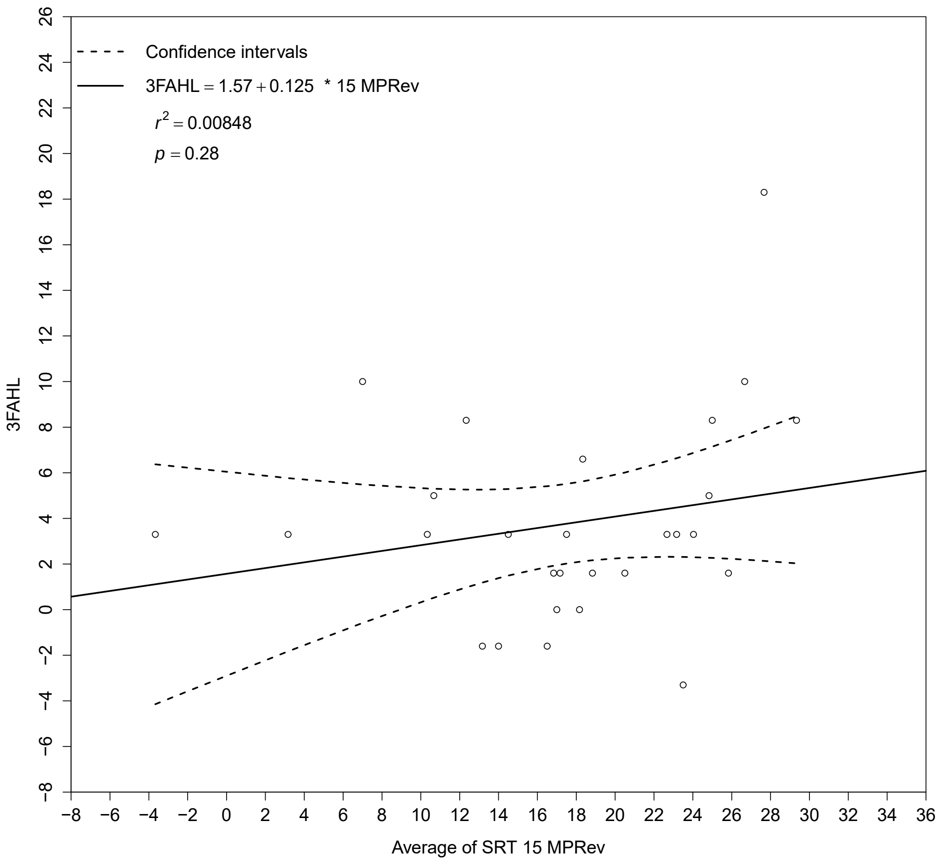


Figure 4.16 Scatter plot for correlation between SRT and 3 FAHL for adults at 15 reversals

4.2.3.2 Results of assessing the SRT with normal hearing children

Reliability

Comparison of speech thresholds at different points of reversals (5, 10 and 15)

Table 4.19 Mean of SRT at 5, 10 and 15 reversals for each age group

	Number of Reversals		
Age group	**5**	**10**	**15**
4 -5 years			
Mean of SRT (dBHL)	18.65	18.90	18.93
Standard Deviation	5.18	5.06	5.15
N	23	23	23
Confidence Interval	2.12	2.07	2.10
Range	16.54-20.77	16.83-20.97	16.83-21.04

Number of Reversals			
Age group	**5**	**10**	**15**
6 -7 years			
Mean of SRT (dBHL)	16.38	17.45	17.37
Standard Deviation	3.43	4.06	3.98
N	21	21	21
Confidence Interval	1.47	1.74	1.70
Range	15.66-19.07	15.72-19.19	15.66-19.07
8 – 9 years			
Mean of SRT (dBHL)	15.59	15.74	15.54
Standard Deviation	4.68	4.30	4.33
N	28	28	28
Confidence Interval	1.74	1.59	1.61
Range	13.85-17.32	14.15-17.34	13.94-17.155
10-13 years			
Mean of SRT (dBHL)	13.06	12.84	12.79
Standard Deviation	4.80	4.26	4.04
N	8	8	8
Confidence Interval	3.33	2.95	2.80
Range	9.74-16.39	9.89-15.80	9.99-15.59

A one-way ANOVA was conducted to compare the mean of SRT across the four age groups at 5, 10 and 15 reversals (Table 4.18). The results showed that there was no significant difference in SRT between the number of reversals in any age group at the $p < .05$ level [$F(2,234) = .164$, $p = .849$] (Figure 4.17).

As shown in Table 4.20 and Figure 4.17, results from the current study show that there are no significant differences for the mean of the SRT between 5, 10 and 15 reversals for children at any age group using INDO-SPRITT material. Thus, this indicates that 5 reversals are enough to estimate SRT with INDO-SPRITT material without sacrificing accuracy and reliability of the thresholds. Although in many cases the accuracy and reliability of the threshold estimate may be improved by obtaining a larger number of reversals, researchers used a different number of reversals to obtain a threshold estimate, and in some studies, researchers used as few as four reversals (Mackie & Dermody, 1986; Bernstein & Gravel, 1990).

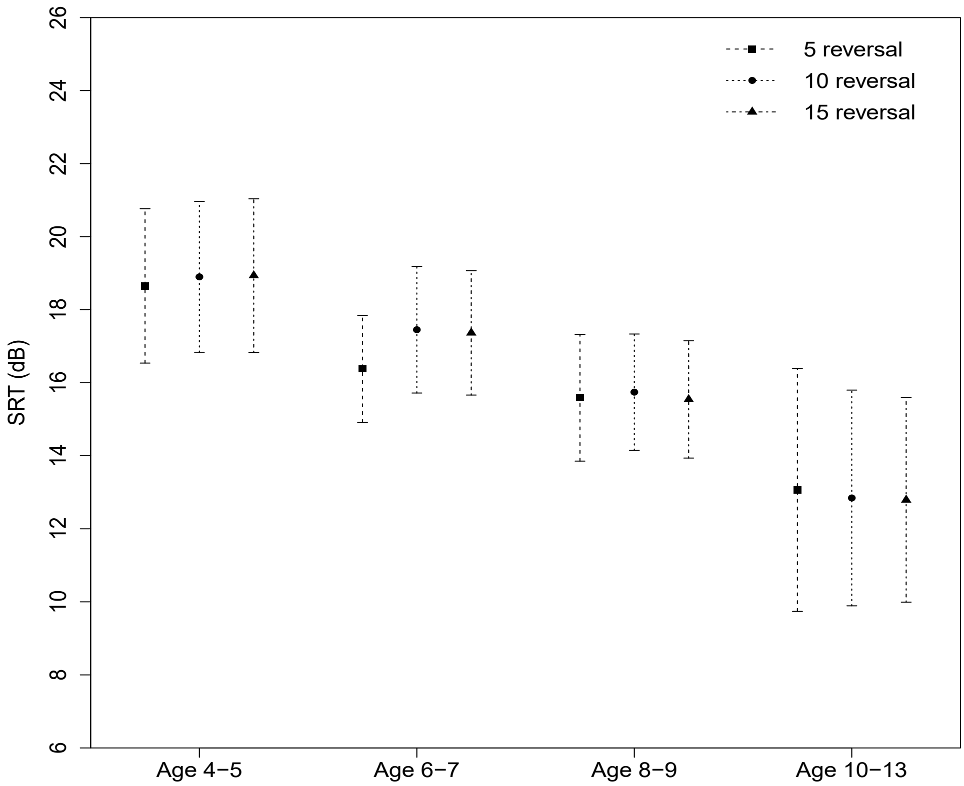


Figure 4.17 SRT for the children across four age groups at 5, 10 and 15 reversals

Having 5 reversals for test administration is an advantage because a fast clinical method is usually more popular with audiologists or hearing health care professionals than a longer method, due to limited times allocated for testing in a clinic. Furthermore, a fast and efficient measurement is important in a child population due to their limited attention span.

As mentioned earlier that the performance of normally hearing individuals provides the normative reference used to identify communication handicap (Doyne, 1951; Soli, 2008), therefore the mean SRTs in the sample study provide the normative reference

(Soli, 2008). The mean that SRTs for children of all age groups at 5 reversals is 16.43 dB, 10 reversals is 16.81 dB and 15 reversals is 16.72 dB (Table 4.20).

Table 4.20 Mean of SRT for all age groups of children

All age groups of children	Reversals		
	5	**10**	**15**
Mean (dBHL)	16.43	16.81	16.72
STDEV	4.78	4.75	4.78
2 STDEV	9.56	9.5	9.56
N	80	80	80
Confidence Interval	1.05	1.04	1.05
Range	15.38-17.47	15.77-17.85	15.67-17.77

As described earlier, 5 reversals are enough to estimate SRT with the INDO-SPRITT material, the normative reference that will be used in clinical practice is SRT = 16.43 ± 9.56 (2STDEV). This gives a range from 6.87 to 25.99. This normative value provides a normative reference against which SRTs of hearing-impaired individuals can be compared.

Thus, elevated SRTs in future clinical practice will provide an objective and quantitative measure of a communication handicap (Soli, 2008). SRTs can then be measured again after medical treatment or the fitting of an hearing aid to give some indication of the extent to which the handicap has been reduced, as well as whether SRTs have improved.

The reliability of a speech perception test refers to ‘the degree to which repeated application of the speech test under identical conditions results in identical scores’ (Dillon & Ching, 1995). This implies that similar conditions need to be specified and maintained during testing and retesting on the basis that consistency of measurements can then be established. In practical terms, however, identical conditions cannot easily be created, because even if the test material is kept identical, that is, a particular speech list is repeated, the participant will be more familiar with the test items when they repeat the test (Dillon & Ching, 1995). Thus, a learning effect may influence the test results.

INDO-SPRITT materials consist of the same words in every list with only the order changed. Also, the same number of foils is used to create equivalent material between lists. Because of this, possible learning effects may be of concern. However, when this type of material is used to determine SRT, learning the stimuli by the subject is unlikely (Dillon & Ching, 1995). For example, in this study an adaptive procedure was used to determine SRT. In this procedure, the intensity is always at a level where the words are just recognisable and because of this, it is harder to learn the stimuli. In other words, a learning effect is unlikely to happen.

In another study by Elliot and Katz (1980), the researchers found that the learning effect did not occur when the same words in every list with different randomisation were presented to a listener in the same test session. This learning effect study was carried out using presentation levels of 0 and 2 dBSL relative to normal hearing children's SRT.

Furthermore, there may be a concern that pictures used in this study may contribute to the learning effect. However, when the children point to a picture associated with the word heard, they are not told whether they are right or wrong, hence their memory is not reinforced by feedback. So, even though they may see the picture several times, it is unlikely that memory will significantly contribute to the learning effect because they were not given feedback. A picture-pointing task was also used by Elliot and Katz (1980) and, as mentioned earlier, a learning effect did not occur in their study.

Figure 4.17 also indicated that there was significant difference in SRT between age groups at the $p < .05$ level [$F(3,234) = 12.882$, $p = .000$].

Further one-way ANOVA was calculated to see if there was a significant difference in SRT values between age groups in each reversals of 5, 10 and 15 (Figure 4.17). Then post hoc test using the Tukey HSD (honest significant difference) test was carried out to determine which groups differ.

The result of a one-way ANOVA for 5 reversal showed that there was a significant difference between age groups at the $p < .05$ level [F(3,76 = 3.598, p = 0.017].

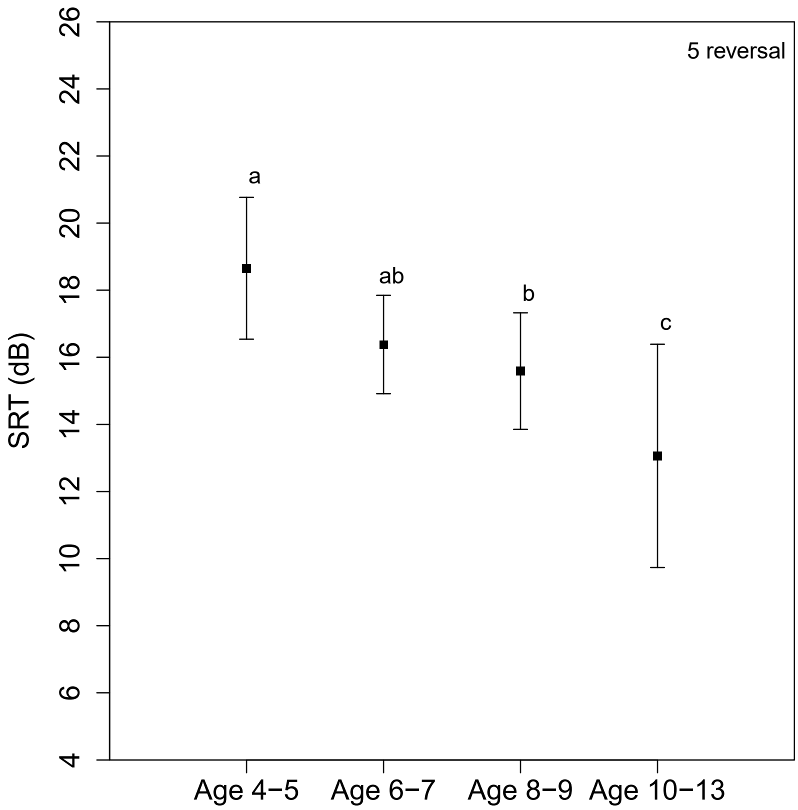


Figure 4.18 Tukey HSD test for age effect at 5 reversals

Where the age groups are labelled (a) there is no significant difference between the SRT data for these groups. The same is true of groups labelled (b). In the case of the ten to thirteen age group, which is labelled (c) the data is significantly different from the other age groups.

The Tukey HSD test for age effect at 5 reversals (Figure 4.18) indicated that:

- The mean score for the 4-5 year old children (M = 18.65, SD = 5.18) was significantly different from the 8-9 year old children (M = 15.5, SD = 4.68), p = 0.019 and the 10-13 year old children (M = 13.06, SD = 4.80), p = 0.029.

- The mean score for 6-7 year old children (M = 16.38, SD = 3.43) did not significantly differ from the age group of 4-5 year old children with p = 0.103 and from age group of 8-9 year old children with p = 0.549. But it was significantly different from the age group of 10-13 year old children with p = 0.029.

- The mean score for 8-9 year old children (*M* = 15.59, *SD* = 4.68) was significantly different from the age group of 10-13 year olds (*M* = 13.06, *SD* = 4.80) with *p* = 0.002 (Figure 4.18).

The result of a one-way ANOVA for 10 reversals showed that there was a significant difference between age groups at $p < .05$ level [$F(3,76) = 4.454$, $p = 0.006$].

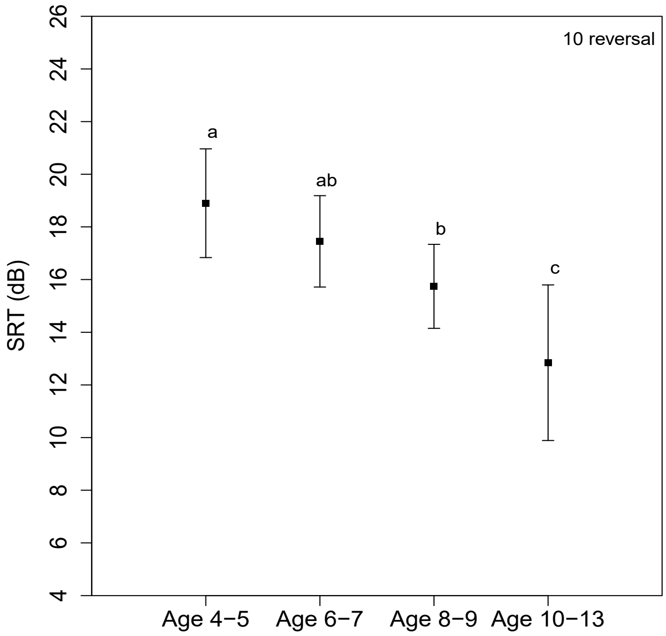


Figure 4.19 Tukey HSD test for age effect at 10 reversals

Where the age groups are labelled (a) there is no significant difference between the SRT data for these groups. The same is true of groups labelled (b). In the case of the ten to thirteen age group, which is labelled (c) the data is significantly different from the other age groups.

The Tukey HSD test for 10 reversals (Figure 4.19) indicated the following:

1. The mean score for 4-5 year old children (*M* = 18.90, *SD* = 5.06) was significantly different from the mean score for 8-9 year olds (*M* = 15.74, *SD* = 4.30), *p* = 0.014 and the mean score for 10-13 year olds (*M* = 12.84, *SD* = 4.26), *p* = 0.043.

2. The mean score for 6-7 year old children (*M* = 17.45, *SD* = 4.06) did not significantly differ from the mean score of either 4-5 year olds (*M* = 18.90, *SD* = 5.06) with p = 0.286 or 8-9 year olds (*M* = 15.74, *SD* = 4.30) with *p* = 0.189. But

it was significantly different from the mean score for 10-13 year olds (*M* = 12.84, *SD* = 4.26) with *p* = 0.015.

3. The mean score for 8-9 year old children (*M* = 17.45, *SD* = 4.06) was significantly different from the 10-13 year olds (*M* = 12.84, *SD* = 4.26), *p* = 0.003 (Figure 4.19).

The result of one-way ANOVA for 15 reversals showed that there was a significant difference between age groups at the $p < .05$ level [$F(3,76) = 4.724$, $p = 0.004$].

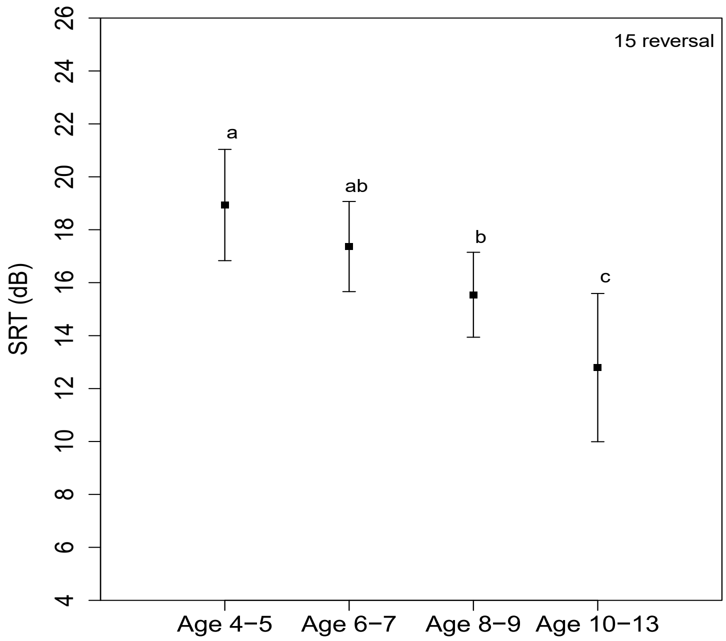


Figure 4.20 Tukey HSD test for age effect at 15 reversal

Where the age groups are labelled (a) there is no significant difference between the SRT data for these groups. The same is true of groups labelled (b). In the case of the ten to thirteen age group, which is labelled (c) the data is significantly different from the other age groups.

The Tukey HSD test for 15 reversals (Figure 4.20) indicated the following:

1. The mean score for 4-5 year old children (*M* = 18.93, *SD* = 5.15) was significantly different from the mean score for 8-9 year olds (*M* = 15.54, *SD* = 4.33), *p* = 0.009 & 10-13 year old (*M* 12.79, *SD* = 4.04), *p* = 0.047.

2. The mean score for 6-7 year old children (*M* = 17.37, *SD* = 3.98) did not significantly differ from the mean score for 4-5 year olds with *p* = 0.249 or 8-9 year olds with *p* = 0.162. But it was significantly different from the mean score for 10-13 year olds (*M* = 12.79, *SD* = 4.04) with *p* = 0.015.

3. The mean score for 8-9 year old children (*M* = 15.54, *SD* = 4.33) was significantly different from the mean score 10-13 year olds (*M* = 12.79, *SD* = 4.04), *p* = 0.003.

The age effect in this study is consistent with other studies. Elliot and Katz (1980) found that at sensation levels lower than 30 dB (0 & 2 dBSL), an age effect was demonstrated, recognising that 10-year-olds performed better than 5-year-olds, who in turn performed better than 3-year–olds. The likely explanation for the age effect observed at low sensation levels is language skill or experience. Mackie and Dermody (1986) found that children could recognise all words when they were presented at a normal conversational level, but when these same words were presented at low speech levels, as in the level used for SRT, the 3-year-olds showed a higher threshold than the 5-year–olds, and the 5-year-olds showed a higher threshold than 7-year-olds. Elliot, Clifton and Servi (1983) found that developmental improvement in performance could be explained by ‘word frequency effects’ which meant children knew and were familiar with the stimulus words although they had less experienced with them compared to adults. Due to these frequency effects, young children (under the age of 8) needed speech to be at a higher intensity level than for older children or adults in order to be understood (Byrne, 1983).

Standard error (SE) measurement based on midpoints of 5, 10 and 15 reversals

Standard error values were calculated for each child’s threshold after 5, 10 and 15 reversals and the mean standard error values for each age group at these reversal points are presented in Table 4.21. Results of the *SE* measurements (Table 4.21) are all less than 2.5 dB for all reversals and age groups. A *SE* of less than half the step size is acceptable (Mackie & Dermody, 1986). The step size in this study was 5 dB. The *SE* decreased in all age groups over 15 reversals. This trend was to be expected, as an increase in the number of reversals should produce a decrease in *SE*.

Random errors of measurement are never completely eliminated and the degree of error in the test consequently needs to be known. A measurement can be labelled as reliable when the amount of random error is slight (Mendel & Danhauer, 1997). The results of *SE* measurements in this study were all less than 2.5 dB for all reversals and age groups. This indicates that INDO-SPRITT is a reliable SRT assessment tool for all reversals and age groups.

Table 4.21 Standard error measurement for children

Number of Reversals			
Age group	**5**	**10**	**15**
4 -5 years			
Mean of SRT (dBHL)	18.65	18.90	18.93
Mean of *SE*	1.32	1.13	0.96
N	22	22	22
Range of *SE*	1.06-1.58	0.98-1.27	0.83-1.09
6 -7 years			
Mean of SRT (dBHL)	16.38	17.45	17.37
Mean of *SE*	0.99	0.92	0.84
N	21	21	21
Range of *SE*	0.79-1.18	0.82-1.02	0.71-0.97
8 – 9 years			
Mean of SRT (dBHL)	15.59	15.74	15.54
Mean of *SE*	1.02	0.89	0.77
N	28	28	28
Range of *SE*	0.76-1.28	0.76- 1.01	0.68-0.85
10-13 years			
Mean of SRT (dBHL)	13.06	12.84	12.79
N	9	9	9
Mean of *SE*	1.30	1.15	0.82
Range of *SE*	0.80-1.81	0.83-1.48	0.61-1.02

Width of the tracking excursion

Finally, the width of the tracking excursion was investigated as an index of test reliability. For each reversal, the excursion width was equal to the dB difference between the upper and lower limits of that excursion. Table 4.22 presents the average excursion width for each age group over the 15 reversals used to calculate SRT. One-way ANOVA was calculated, resulting in a significant difference between age groups at $p < .05$ level [$F(4,56) = 573.4$, $p < 0.01$]. Tukey's HSD test was then computed and the results showed that the mean of excursion width of 4-5 year olds ($M = 5.72$, $SD = 5.75$) was significantly different from the mean excursion width of 10-13 year old children ($M = 6.63$, $SD = 5.16$) with $p = 0.0196$.

Note that the number of observations ($N = 150$) in the 10-13 year age group was much smaller than the rest of the groups, which may have affected the results. The means for the rest of the group comparisons were not significantly different from one to another ($p > 0.05$).

Table 4.22 Average excursion width on the INDO-SPRITT for each age group over 15 reversals

	Age Group			
	4-5 Years	**6-7 Years**	**8-9 Years**	**10-13 Years**
Average	5.72	6.29	6.33	6.63
Standard Deviation	5.75	5.43	5.45	5.16
N	510	420	570	150
Standard Error	0.25	0.26	0.23	0.42
Confidence Interval	0.50	0.52	0.45	0.83
Range	5.22-6.22	5.77-6.81	5.89-6.78	5.81-7.46

Overall, the results were similar for each age group. Given the step size of 5 dB used in this study, each age group on average only required one word above or below the SRT point to establish a reversal (Figure 4.21).

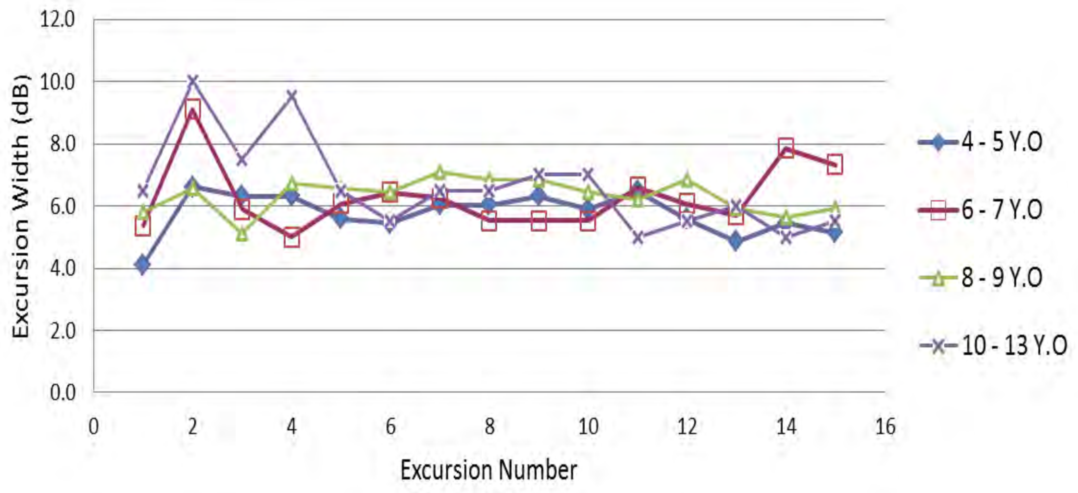


Figure 4.21 Mean excursion width on INDO-SPRITT for each age group over 15 reversals

In view of the results of all reliability measurements, it can be concluded that INDO-SPRITT is a reliable assessment tool for measuring SRT.

Validity

Predictive validity

A Pearson product-moment correlation coefficient was computed to assess the relationship between SRT and the 3 frequency average hearing level (3 FAHL) (500 Hz, 1000 Hz and 2000 Hz). There was a positive correlation between the two variables as shown in Table 4.23.

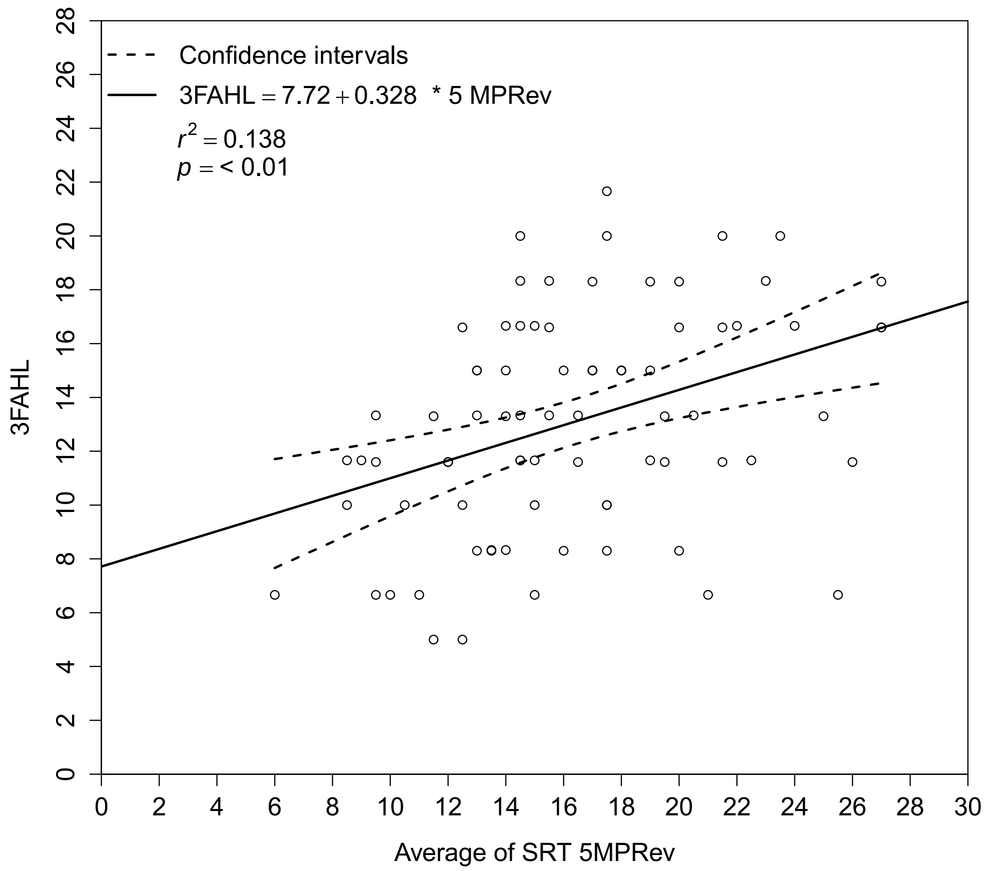
Table 4.23 Mean SRT and 3 FAHL of all age groups

All age groups of children	**Reversals**			**3 FAHL**
	5	**10**	**15**	
Mean	16.43	16.81	16.72	13.11
STDEV	4.78	4.75	4.78	4.06
2 STDEV	9.56	9.5	9.56	8.12
N	80	80	80	80
CI	1.05	1.04	1.05	.0.89
Range	15.38-17.47	15.77-17.85	15.67-17.77	12.22-14.00

The linear regression plots summarise the results (Figures 4.22 & 4.23). Overall, there was a modest, positive correlation between SRT and 3 FAHL as increases in SRTs were correlated with increases in 3 FAHL (Table 4.24).

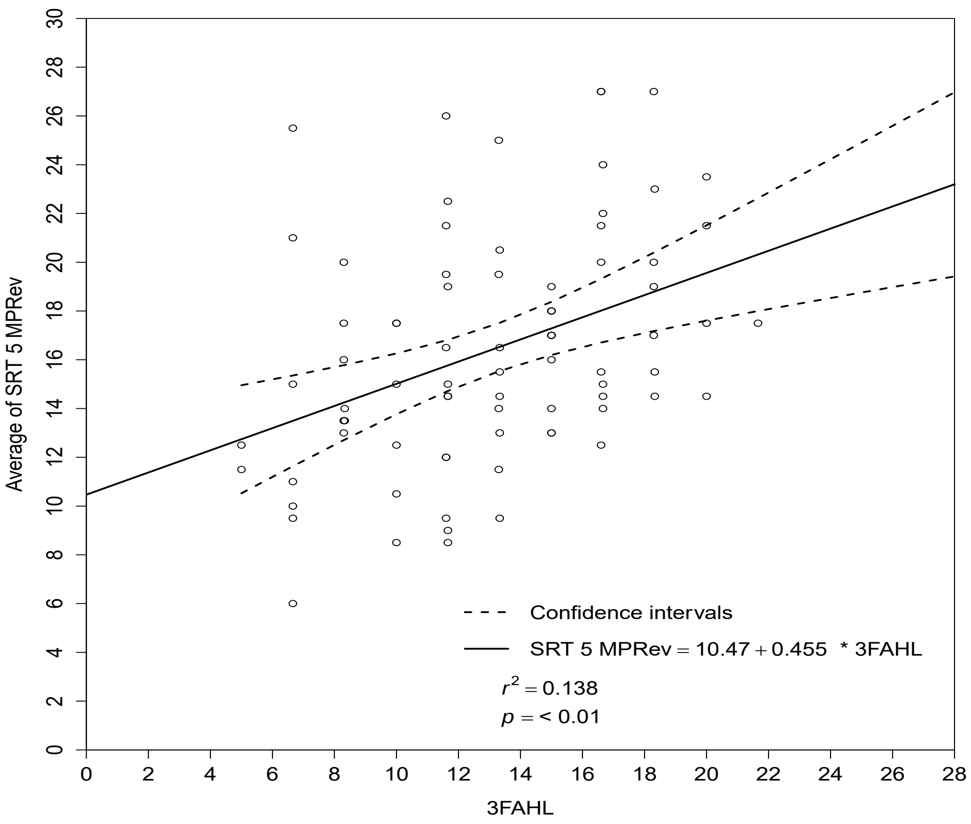
Table 4.24 Correlation between SRT and 3 FAHL

Reversals	Correlation between SRT and 3 FAHL
5	$r = 0.386, p = 0.000$
10	$r = 0.489, p = 0.000$
15	$r = 0.471, p = 0.000$



Note: 5MPRev: 5 mid points of reversals (5 reversals)

Figure 4.22 Linear regression plot between SRT as the predictor and 3 FAHL as the criterion at 5 reversals for children



Note: 5MPRev: 5 mid points of reversals (5 reversals)

Figure 4.23 Linear regression plot between 3 FAHL as the predictor and SRT as the criterion at 5 reversals for children

As mentioned earlier in this study, 5 reversals are enough to estimate SRT, therefore, the formula that will be used in future clinical practice to predict 3 FAHL (Figure 4.22) and SRT (Figure 4.23) for hearing ≤ 20 dBHL (as the subjects in this study were children with hearing ≤ 20 dBHL), is as follows:

To predict 3 FAHL, the formula is : 3 FAHL (in dBHL) = 7.72 + (0.328 x SRT)

95% confidence interval for the regression slope to predict 3 FAHL is 0.152 to 0.505

To predict SRT, the formula is : SRT (in dBHL) = 10.47 + (0.456 x 3 FAHL)

95% confidence interval for the regression slope to predict SRT is 0.210 to 0.699

For clinical purposes it might be more appropriate to use figures rounded up as follows:

$$3\ \text{FAHL (in dBHL)} = 8 + (0.33 \times \text{SRT})$$
$$\text{SRT (in dBHL)} = 10.5 + (0.5 \times 3\ \text{FAHL})$$

The procedures and formulae used above are from the SPSS statistical package.

The accuracy of this estimate for each child in the clinical practice will be ± 2 standard deviations (2*SD*). The 2*SD* for 5 reversals is 9.56 (Table 4.23). Thus the formula to predict SRT and 3 FAHL, rounding up the values becomes:

$$3\ \text{FAHL (in dBHL)} = 8 + (0.33 \times \text{SRT}) \pm 10$$
$$\text{SRT (in dBHL)} = 10.5 + (0.5 \times 3\ \text{FAHL}) \pm 10$$

The generally accepted value for good agreement between SRT and PTA is that they are within ± 6 dB of each other (Brandy, 2002; Olsen & Matkin, 1991; Carhart, 1971). However, many studies on the relationship of SRT to PTA used spondaic words and were carried out in the English language.

Mackie and Dermody (1986) undertook a study to find the relationship between SRT and pure tone hearing level by using monosyllabic words from the NU-CHIPS material. They did not find any correlation between SRT and 3 FAHL with subjects who had normal hearing but found a significant correlation with hearing loss ≥20 dBHL and the formula they arrived at was as follows (Lovegrove, Dillon & Mackie, 1991):

$$\text{SRT (in dBHL)} = 3\ \text{FAHL} \pm 10\ \text{dB}$$

This would seem to indicate that different speech material can produce different relationships between SRT and PTA. As mentioned earlier, the relationship between speech reception thresholds and PTA is dependent on several factors, such as method, configuration of the loss, type of threshold obtained and speech materials used (Kruger & Mazor, 1987). Thus, one cannot make the assumption that values similar to those found when using English speech material would automatically be applicable to the Indonesian language without undertaking studies on this matter.

An attempt was made to investigate how the English formulae were arrived at. However, there seem to be little information on this matter.

Face validity

INDO-SPRITT has good face validity because of its use of speech stimuli for hearing tests. This has higher face validity than those tests that use pure tones. Speech is thus easier to justify as a test stimulus and makes sense to the layperson (Henkel, 2005) as most human communication is through speech. Furthermore, INDO-SPRITT uses speech material that is familiar to children because if the words used are not familiar, the children are less likely to recognise them and poorer scores will result.

Content validity

INDO-SPRITT has good content validity because of its use of speech material that meets the criteria for content validity, which is phonemically balanced, uses lists of equal difficulty and uses familiar vocabulary (Walden, 1984). Thus, part of content validity and familiar vocabulary contributes to face validity. As mentioned earlier, if content validity has been appropriately determined, then face validity will also be appropriate (Walden, 1984).

As mentioned earlier, the INDO-SPRITT material has lists of equal difficulty because it consists of the same words in every list with only the order changed, as well as the same number of foils in each list.

Construct validity

INDO-SPRITT has good construct validity because INDO-SPRITT material uses meaningful bisyllabic words that are age appropriate to the target population and therefore it can provide information, to some extent, on speech perception in communication (in quiet environment). According to Mackie and Dermody (1982), a speech test would be considered to have construct validity if it provides information on speech perception in communication and the circumstances influencing it.

Furthermore, speech perception is an abstract construct and no single observable parameter can represent the abstract construct of speech perception exactly (Mendel, 2008; Walden, 1984). Therefore, a test battery is needed to measure specific aspects of speech recognition, such as phoneme recognition, word recognition, sentences spoken by different talkers and various conditions. It is intended to develop further speech perception assessments for the Indonesian population in the near future.

4.2.4 Future studies

It is hoped that the use of INDO-SPRITT in the future would help hearing-impaired children by increasing the awareness of parents, teachers and the general public on the implications of deafness. As awareness increases, varying degrees of hearing loss may be better detected and recorded. Therefore, further studies with INDO-SPRITT, involving its standardisation with hearing-impaired subjects, could be carried out. In addition to this, further development of INDO-SPRITT in noise environments should also be considered. If facilities for free field-testing in hearing clinics in Indonesia are adequate, SRT assessments in a free field setting can be carried out. A free field presentation may enable children's parents or teachers to observe at first hand the effects of hearing loss and amplification on the child's speech reception.

A further study that may be required is to investigate factors that may increase the magnitude of the correlation between 3 FAHL and SRT of child subjects. Furthermore, if INDO-SPRITT is to be used for the adult population, a further study that includes a larger sample should be carried out to standardise the test and to provide a normative reference for this population.

CHAPTER 5:
INTRODUCTION TO INDONESIAN SPEECH PERCEPTION ASSESSMENT FOR SEVERELY-PROFOUNDLY HEARING-IMPAIRED CHILDREN (INDO-SPASP)

Much is already known about the effect of hearing impairment on speech perception at the phonetic level (Erber, 1972, 1974, 1981; Pickett et al., 1972; Walden & Montgomery 1975; Bilger & Wang, 1976; Risberg 1976; Hack & Erber, 1982). For example, we know that speech perception performance, on average, decreases with increasing hearing loss. We also know that for a given degree of hearing loss some aspects of speech are perceived auditorily better than others. Suprasegmental patterns, for example, are perceived better than segments, vowels better than consonants, vowel height better than vowel place, consonant voicing better than consonant manner, and consonant place the least accurately (Boothroyd, 1984).

The term profoundly hearing impaired is commonly applied to children with unaided pure tone 3-frequency averages of 90 dB or greater (Moog & Geers, 1990; Plant, 1995). Erber (1979) suggested that profoundly hearing-impaired children can best be characterised by those whose ability to distinguish spectral features in speech is extremely poor, and for whom the gross variations in the acoustic pattern (waveform envelope) are the principal information available for speech processing. Thus, in general, it is the suprasegmental aspects or prosody of speech that are likely to be most accessible to profoundly deaf children as also mentioned by Maltby (2000). Martony (1974) mentioned that deficiencies in perception of signal durations were also observed in profoundly hearing impaired children. However, the ability to perceive small differences in intensity is not so greatly impaired in profound hearing impairment.

Many profoundly deaf children score very poorly on traditional speech perception test procedures (Markides, 1980; Dowell & Cowan, 1994), such as Northwestern University Children's Perception of Speech (NU-CHIPS) by Elliot and Katz (1980), and Word Intelligibility by Picture Identification (WIPI) by Ross and Lermann (1970). A score of zero on a conventional speech perception test, however, may incorrectly suggest that a profoundly deaf child is unable to perceive any auditory information. It is important to know how much use these children can make of their residual auditory acuity and how sensitive they are to the differences in speech patterns that are used to define word meanings and sentence structures via hearing aids (Boothroyd, 1989).

As previously discussed, during a field study in Indonesia the subjects were found to have severe to profound hearing loss and the design of INDO-SPRITT is not suitable for this population (Section 4.2.2), as they could not achieve any meaningful scores.

As a result of this experience, the need for a speech perception test specifically designed for the profoundly hearing impaired in Indonesian Language become obvious. Testing procedures designed for severely-profoundly hearing impaired children need first to address themselves to differentiating between those children who can perceive the spectral/segmental components and those who can perceive only time intensity patterns/ suprasegmental components (Plant, 1984).

The information obtained from such a speech perception test will be valuable in the child's habilitation, whether from an educational point of view or possibly from the perspective of the choice of a hearing device. Often, to suggest an educational placement or to plan the child's auditory training, a clinician must establish whether a young hearing-impaired child perceives the spectral qualities of sounds or seems to perceive only their intensity patterns. Furthermore, although profoundly deaf children are a difficult group to aid successfully, aiding residual hearing, however minimal, is an important part of support for such children. Successful aiding is vital in the development of listening skills

Furthermore, children with severe to profound hearing impairment demonstrate a range of speech perception abilities (Cramer & Erber, 1974; Erber, 1974; Risberg et al., 1975; Erber, 1979, 1980). Therefore, a speech perception test with a hierarchical order of difficulty is important.

As in many developing countries, Indonesia lacks resources for the diagnosis and rehabilitation of hearing impairment/deafness and the skill levels of the available human resources such as teachers of the deaf, speech therapists, audiologists and audiometricians is less than optimal because of a variety of reasons attributable to training, deployment and work environment (WHO, 2004). Therefore, the purpose of this study is to provide a measure of speech perception capacity in subjects with limited hearing which can easily be applied in many underdeveloped areas in Indonesia. Thus, the test should require minimal investment in technology, be very easy to administer, and have a variety of pragmatic applications.

There are many speech perception tests for profoundly deaf children available in English, such as spondee recognition tests (Cramer & Erber, 1974; Erber, 1974); Children's Auditory Test (CAT) (Erber & Alencewicz, 1976), which was later expanded to the Monosyllable, Trochee, Spondee test (MTS or MonSTr) (Erber & Witt, 1977); Auditory Numbers Test (ANT) (Erber, 1980); Merklein Test (Merklein, 1981); Glendonald Auditory Screening Procedure (GASP!) (Erber, 1982, 2011); Plant and Wescott Test (PLOTT) (Plant & Westcott, 1983); Three-Interval Forced Choiced test of speech pattern contrast perception (THRIFT) (Boothroyd, 1991); Early Speech Perception (ESP) (Moog & Geers, 1990) and Maltby Speech Perception Test (Maltby, 2000).

It has been suggested that THRIFT would be an ideal test to be adopted as it is said to be language independent. THRIFT (Boothroyd, 1991) was developed to provide a measure of speech perception capacity in subjects with limited hearing and it was intended to be language independent as the client group in New York included children with a variety of linguistic background. It has two purposes:

1. To generate a profile of an individual's access to several phonologically significant speech pattern contrasts.

2. To provide a single numerical value of merit. This means, the test is intended to measure the amount of sensory evidence that the subject obtains from speech, regardless of current age, age at onset of deafness, listening experience, motor speech skill, and language development

This test evolved from work with the Speech Pattern Contrast test (SPAC) (Boothroyd, 1988), the design of which was based on Risberg's rhyme test (Risberg, 1976). THRIFT was designed to avoid the need for vocabulary knowledge and reading skills. It uses an oddity task. Each stimulus consists of three utterances. Two are repetitions of the same syllable (CV, VC or CVC. C: consonant and V: vowel). One differs from the other two along a single phonologically significant dimension of one of its components. In 'taw taw daw', for example, the difference is in the voicing of the initial consonant. The subject's task is to decide whether the odd item was first, second, or third in the sequence of three utterances. Responses can be verbal, by pointing, or by key press.

There are nine subtests of THRIFT test:

1. Intonation (e.g. 'peh?' versus 'peh!')
2. Vowel height (e.g. 'saw' versus 'sue')
3. Vowel place (e.g. 'doo versus 'dee')
4. Initial consonant voicing (e.g. 'taw' versus 'daw')
5. Final consonant voicing (e.g. 'eez' versus 'eece')
6. Initial consonant continuance (e.g. 'seh' versus 'teh')
7. Final consonant continuance (e.g. 'awz' versus 'awd')
8. Initial consonant place (e.g. 'foo' versus 'soo')
9. Final consonant place (e.g. 'eeg' versus 'eed')

The nine subtests consist of one suprasegmental contrast (Subtest 1) and eight segmental contrasts (Subtests 2-8). The eight segmental contrasts consist of two vowel contrasts and six consonant contrasts. However, the balance of eight segmental contrasts (Subtests 2-9) to one suprasegmental contrast (Subtest 1), probably underestimates the importance of the latter. It is also unclear as to why THRIFT emphasises intonation only for the suprasegmental component.

As mentioned above, testing procedures designed for severely-profoundly hearing impaired children need first to address differentiating between those children who can perceive the spectral/segmental components and those who can perceive only time intensity patterns/ suprasegmental components (Plant, 1984). In Indonesia, such a test does not yet exist.

Therefore, at the initial stage of the development of the INDO-SPASP battery, a subtest which distinguished between those children who can perceive the spectral/segmental components and those who can perceive only suprasegmental components was developed. Tests such as a spondee recognition test, CAT, MTS or MonSTr and ANT are spectral/suprasegmental tests that can be useful in making such diagnostic decisions. They were all designed to evaluate the profoundly hearing-impaired child's ability to perceive durational and stress patterns in speech and to discriminate among words mainly on the basis of vowel sounds. Therefore, at this initial stage, these tests developed by Erber served as the basis for a subtest battery of INDO-SPASP instead of using THRIFT.

Another reason for not using THRIFT is that in the English language, many words are monosyllabic and most of them consist of a CVC syllable pattern. Thus, the balance of two vowel contrasts to six consonant contrasts in THRIFT is probably appropriate. However, unlike English, most BI words are bisyllabic, possessing the following patterns (Dardjowidjojo, 1978):

1. V I-kan, u-dang
2. VC in-dah, um-pan
3. CV ru-mah, bu-ku
4. CVC kun-ci, pan-tai

Thus, the composition of vowels in the syllable pattern of BI is different to an English syllable pattern and the balance of two vowel contrasts to six consonant contrasts is probably not appropriate for the Indonesian language.

Even though THRIFT was designed to avoid the need for vocabulary knowledge and reading skills, it is possibly not applicable for use in Indonesia because the test is based on English speech patterns. One might think that this type of test should be easily adapted to other languages due to its vocabulary independence. Whilst THRIFT is an extremely useful test in some contexts, it is not language independent; it is only vocabulary independent with regard to the English language or those languages with similar characteristics.

In summary, the main rationale of not using THRIFT, at this stage, is that it is lacking in suprasegmental assessment. The second reason is that, even though THRIFT was designed to avoid the need for vocabulary knowledge and reading skills, it is not as easily applicable to the Indonesian population as one might think, as the test is based on a common English speech pattern.

Proposed uses of INDO-SPASP are as follows:

1. To assist in educational placement. Since a child's perception ability and educational placement/environment would impact each other, along with other factors, the speech perception abilities of the child might provide an indication of the most appropriate educational methodology. For example, auditory-verbal (auditory alone), auditory-oral (auditory visual) or total communication (auditory, visual, sign language, and finger spelling). The criteria for the educational placement might take into account the child' segmental perception ability as follows:

    - Segmental score ≥80% may be placed in an auditory-verbal class.

    - Segmental score between 50% and 80% may be placed in an auditory-oral class.

    - Segmental score 50% may be placed in a total communication class (Moore, 2014)

2. To establish the objectives for auditory training. If the INDOSPASP test battery shows that the child is unable to perceive differences in common speech patterns, the objective should be to get the child to differentiate among patterns with gross

differences. If the child is able to discriminate two syllable words from four syllable words, but not two syllables of long single waveform from two syllables of double waveforms, the objective might be to refine the child's pattern discrimination skills. If the child is beginning to differentiate among some words of similar durational and rhythmic pattern, the training objective would be to increase this ability.

3. To measure the effects of auditory training. The INDO-SPASP battery can be used to measure progress in auditory skills over time, no matter what listening device the child is using. There are four levels of auditory skill with respect to speech perception: detection/awareness, discrimination, identification/recognition and comprehension (Erber, 1982, 2011). This study, at this stage, only involves phoneme detection as Subtest 1 and identification of words as Subtest 2. (It is hoped that further subtests to assess higher levels of auditory skill will be developed in the future.) Administering the test on a regular basis, such as at the beginning and end of a school year or every six months in a clinical setting, will make it possible to document any changes in the child's ability to perceive speech.

4. To assist in selecting a listening device for the child.

5.1 Subtest 1 – Phoneme detection

A detection test is included in this test battery as it is the most basic auditory skill (Tye-Murray, 2009). It is the basic process of determining whether sound is present or absent (Small, 1973; Erber, 1982, 2011). The detection subtest can be used for several purposes:

1. To determine if the child's hearing aid is minimally adequate for auditory speech perception, providing enough gain and output level so that most phonemes can at least be detected.

2. To establish whether the teacher's voice is heard by the child at a sufficiently high level for auditory communication.

3. To establish whether it is reasonable to present the second subtest. This avoids the situation where the person administering the test is attempting to test a child's speech perception abilities with inaudible sound.

Some tests such as PLOTT and MALTBY use monosyllabic words as a detection test. INDO-SPASP, however, will use phonemes for this purpose. It is easier to develop a detection test using phonemes as opposed to using monosyllabic words due to the nature of the Indonesian language in which monosyllabic words are uncommon. This subtest will follow the Ling Sound Test (Ling, 1976) as the basis of a detection test. The concept behind Ling's sound test was to select familiar speech sounds that would broadly represent the speech spectrum from 250 Hz to 8000 Hz. This spectral range is the same frequency range tested by conventional audiometry. Ling used isolated phonemes to target low, middle and high frequency sounds.

The rationale for using the Ling Sound Test as part of this test battery is that it is widely recognised that some of its benefits are that it can be used by anyone; audiologists, speech language pathologists, teachers and parents. It can be used with hearing aids, cochlear implants, or no amplification at all and it covers the speech range from low frequency to high frequency and from voiced vowels to unvoiced consonants. It has been used as a good low technology tool for quick and accurate assessment of the essential communication abilities of adults and children and it is very easy to learn. Therefore, the Ling Sound Test should be highly applicable in Indonesia as audiology is still underdeveloped and many areas in Indonesia do not have highly skilled professionals and are not well equipped with high technology. Furthermore, the Ling Sound Test is also useful for addressing the higher auditory skills such as discrimination and identification (but not as a test of comprehension). Hence, the Ling Sound Test itself can be used as an independent tool to assess the hierarchy of auditory skills and this should be very useful for Indonesia. A separate section (see page 121 to 124) will discuss further use of the Ling Sound Test for Indonesia.

The Ling Sound Test was originally developed for the North American population. Therefore, there might be differences in production and spectral content of North

American and Indonesian vowels and consequently this chapter will discuss the appropriateness of the Ling sounds for an Indonesian population.

The Ling Sound Test was developed by Ling (1976) as a quick, easily administered test with high face validity for verifying the effectiveness of hearing aid fitting in children. The original Ling test employed five speech sounds spanning the speech frequency range, commonly referred to as the Ling five sounds, which include: the fricatives /ʃ/ and /s/ as in ‘fish’ and ‘us’ respectively, as well as the corner vowels /i/ as in she, /u/ as in two and /a/ as in car (Ling, 1976). The bilabial stop consonant was added by Ling et al. (1995) due to its low frequency content and the test was subsequently called the Ling six sound test. The Ling Sound Test has been adapted for an Australian population and the use of six sounds has previously been adjusted by replacing /u/ with /ɔ/ due to the differences in production and spectral content (Agung et al., 2005) (Figure 5.2).

The six sounds were selected for a number of reasons. The nasal consonant/m/ has a formant structure similar to that of vowels but with nasal formants at around 250, 2500 and 3250 Hz (Ladefoged, 1993; Ladefoged & Johnson, 2011). The higher formants are reduced in intensity and therefore the peak energy of /m/ occurs at a low frequency of approximately 250-350 Hz (Australian Hearing Manual of Speech Perception, 1999). Inability to hear /m/ at a distance of approximately three metres indicates that a child is unlikely to adequately hear other low-frequency sounds that are important for the development of normal speech. As a result, nasalised speech and abnormal prosody may occur (Ling, 2002). Therefore, the sound /m/ is used in the Ling six sound test to verify whether hearing for low frequencies is satisfactory.

The voiceless fricatives /ʃ/ and /s/ (also known as sibilants) have little energy below their high-pass cut-off frequency, which for /ʃ/ is approximately 2000 Hz and for /s/ is approximately 3500 Hz. The ability to detect these sounds can be used to determine whether mid-high frequency sound is audible to a hearing–impaired child while wearing amplification (Ling, 2002). Audibility of higher frequency sounds is important because they carry key linguistic cues, for instance, /s/ is commonly used in English to signal plurality and tense. Therefore, reduced audibility of /s/ may result in delayed development of linguistic rules (Stelmachowicz et al., 2001). It is also important to note that although both /s/ and /ʃ/ have little energy below their high-pass cut off, the dominant spectral energy is at a higher frequency for /s/ (around 5000 Hz) than for /ʃ/ which has dominant spectral energy at about 3000 Hz (Harrington & Cassidy, 1999).

Ling (1976) devised the Ling six sound test using the corner vowels: /i/ as in she, /u/ as in two, and /a/ as in car. These corner vowels are the most common vowels, and are present in a large majority of the world's languages (Boe et al., 2002). The ability of a listener to easily identify any given vowel is primarily dependent upon the audibility of the first formant (F1) and second formant (F2) of the vowel. In American English, /a/ is a low central vowel (F1 and F2 are located at the centre of the vowel frequency range), /u/ is a high back vowel (F1 and F2 are low frequency) and the vowel /i/ is a high front vowel (F1 is low and F2 is high frequency).

The vowel /a/ is used to check that central vowels are audible and not under- or over amplified. If a child has reasonably good low-frequency hearing and can detect /a/ with their hearing aids on, but not low frequency sounds such as /m/, this suggests that either low frequency sounds are under amplified or sounds in the mid frequency range are over amplified. According to Ling (2002), grammatical errors in a child's speech may result if the child cannot detect /a/, because less intense, unstressed words, particularly those in the centre of the speech range, are likely to be missed, for example, 'and' [ənd], in connected speech.

The vowel /u/ in American English, is a high back vowel (F1 and F2 are low frequency). According to Ling (2002), the vowel /u/ should be used to check whether the low end of the vowel formant range can be detected, as F1 and F2 have low formant frequencies. /u/ is also used to ensure that low frequency sounds such as/u/ and /m/ are recognised and not confused with each other.

The vowel /i/ may be detected through the audibility of either the low first formant or the high second formant, which is at the upper end of the frequency range of vowels. Thus, children may respond to /i/ because they detect the low formant only or the high formant only or both. Therefore, the interpretation of /i/ is the most complex of all. Their response to /i/ must be interpreted by comparing it to their response to other sounds in the test that have components in the same frequency range, as follows;

1. No response indicates:
   i) Neither formant can be detected.
   ii) Interpretation of Ling results: Poor low-mid/high frequency hearing.

2. A response could indicate three possibilities:
   i) Detection of only low frequency formant if [u] and [m] can be detected but [ʃ] and [s] cannot. Interpretation of Ling results: good low frequency hearing, poor mid-high frequency hearing.
   ii) Detection of only the high frequency formant if [ʃ] can be detected but [m] and [u] cannot. Interpretation of Ling results: good mid- high frequency hearing, poor low frequency hearing.
   iii) Detection of both formants, if the [ʃ], [s], [u] and [m] can also be detected. Interpretation of Ling results: good hearing up to about 2500 Hz.

Table 5.1 Six vowels of BI

	Unrounded		Rounded
	Front	Central	Back
High	i		u
Mid	e	ə	o
Low		a	

Because Ling (1976) devised the Ling six sound test using the corner vowels for a North American population, differences between the American and Indonesian language vowels need to be addressed when this test is used clinically. As mentioned in the previous chapter, there are six monophthong vowels in the Indonesian language (Lapoliwa, 1981) and for the Ling test, only corner vowels are used.

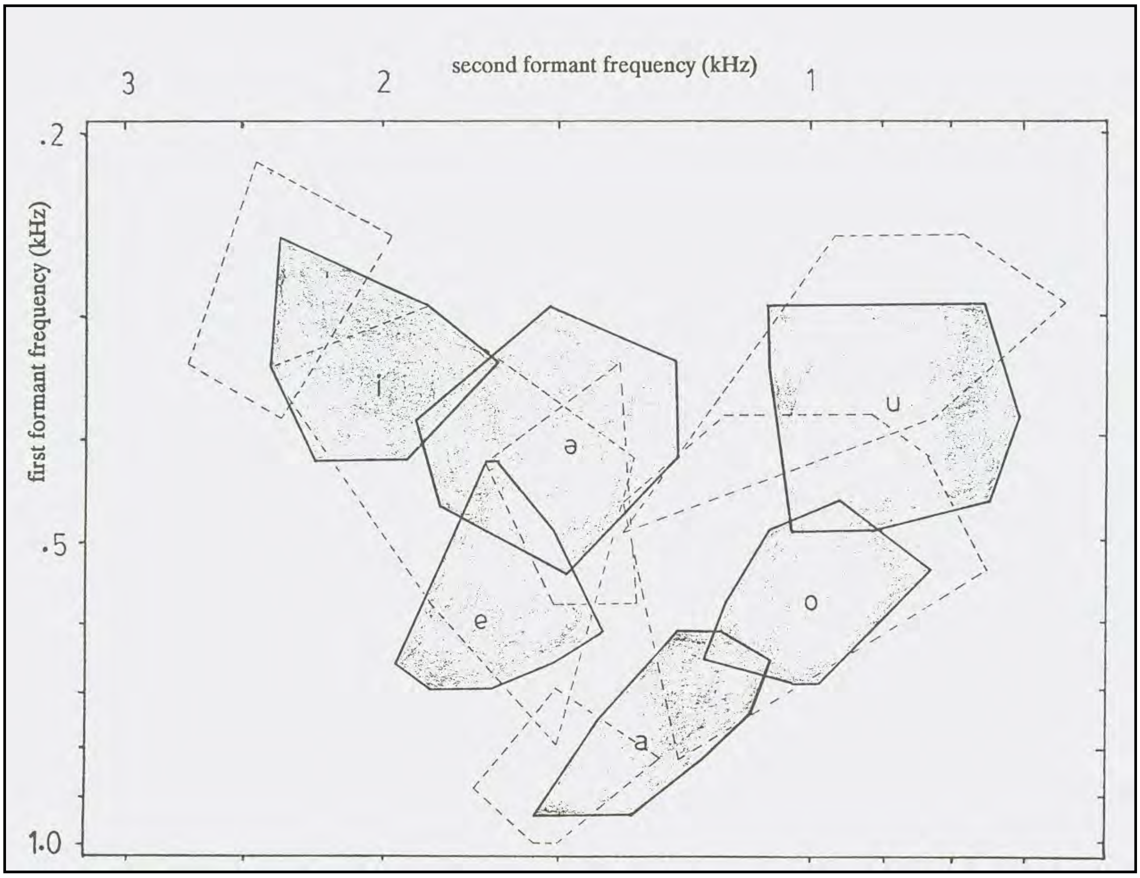


Figure 5.1 Indonesian vowel formants in context and in isolation

Scatter diagram of the six vowels spoken in context and in isolation. The shaded polygons with their continuous straight demarcation lines from the envelopes of the 50 realisations per vowel in context (10 speakers, five realisations per speaker). The centres of gravity (i.e. the intersections of F1 and F2 values of the vowels concerned) are indicated with the symbols of the phoneme concerned; F1 and F2 values are averaged first over utterances per speaker and then over the speakers. The polygons with dashed demarcation lines envelop the 10 tokens of each of the vowels spoken in isolation (10 speakers, one token per speaker).

This table was taken from Ellen van Zanten (1989), The Indonesian vowels; Acoustic and perceptual explorations. With permission from Ellen van Zanten, Leyden University, Netherlands.

The comparison of Indonesian, American and Australian vowel formants are shown in Figure 5.2. It is clear that the Indonesian corner vowels are more similar to the American corner vowels compared to the Australian corner vowels. In Australian English the corner vowel /u/ is high central. For this reason the high back vowel for Australian English /ɔ/ is used for testing children in Australian English.

Table 5.2 Formant structure for vowels in isolation (Ellen van Zanten, 1989)

	F1		F2		F3	
	Mean	**STDEV**	**Mean**	**STDEV**	**Mean**	**STDEV**
/i/	287	50	2413	248	2860	365
/e/	459	139	1730	341	2435	266
/a/	877	92	1529	126	2447	297
/o/	471	137	1015	208	2263	485
/u/	320	75	914	197	2406	261
/ə/	486	87	1493	112	2647	195

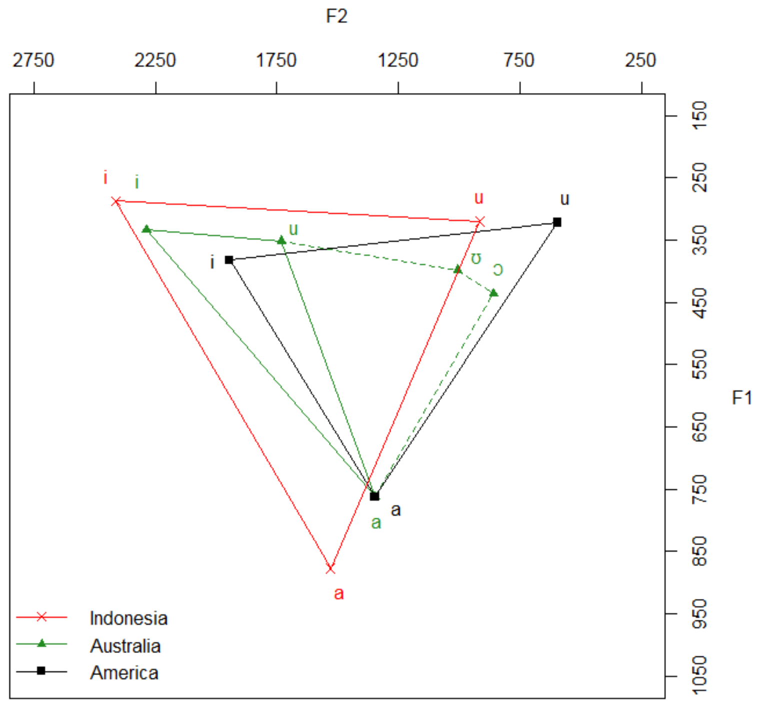


Figure 5.2 Comparison of American, Australian & Indonesian vowels

American English versus Australian English versus Indonesian vowel formants. American data from Ladefoged (1993); Australian data from Cox (1999, 2012); Indonesian data from Van Zanten (1989).

The location of vowel /ɔ/ is similar to the American /u/. It can be seen in Figure 5.2 that the /ɔ/ vowel is the most high and back long vowel of Australian English. Indonesian /u/ has a very clear high back pronunciation similar to American /u/. Figure 5.2 indicates mean value of these vowels, but the actual individual Indonesian vowels spread around this point similar to that on Figure 5.1. The Indonesian /u/ vowel would have similar distribution (vowel space) to that of American English.

Using the Indonesian vowel space in this manner as in Figure 5.1 allows us to use vowel formant frequencies as stimuli which indicate a child ability to perceive sounds at different points in the speech frequency range.

In the absence of a study outlining an acoustic comparison of Indonesian and English consonants, it is possible to make some perceptual judgements about the similarity of some consonants. If Indonesian /s/ and English /s/ both sound similar, then this would be due to similarity between these sounds in the two languages. That is, in both languages they will have similar high frequency characteristics. English speakers easily recognise the /s/ in Indonesian and vice versa, because the frequency of this consonant is very similar for the two languages. Also the /ʃ / sounds sound very similar between the two languages. Also, these two sounds are very common in both languages. This means that in both languages ability to perceive these two consonants indicates that the listener can hear frequency of 3000 Hz and above. In English, the /s/ sound has a high pass cut-off at 3000 Hz which spreads above that frequency for male speakers, and is about 5000 Hz for adult female speakers, and even higher for children. In English, the /sh/ sound has a high frequency cut-off is at about 2000 Hz and above for male speakers, and is about 3000 Hz for adult female speakers, and even higher for children. (Harrington & Cassidy, 1999). Ability to hear these sounds would imply that the listener is able to hear these higher frequencies and inability to hear these sounds would imply that the listener cannot reliably hear these higher frequencies.

Thus, unlike the Australian version of Ling Sound Test, it would seem that modification of the Ling Six sound test is not necessary for its use with the Indonesian population.

This feature of the Ling Six Sound test, where the use of speech stimuli can give an indication of the audibility of various frequencies is of great importance for those helping children to acquire oral language. Thus, the philosophy behind the detection test using phonemes is to determine whether certain frequencies are perceived and does not focus specifically on the identification of vowels or consonants (Ling, 1976).

5.2 Subtest 2 – Identification of words

Identification (*Recognition*) is the ability to label or name the sound that is heard (Erber, 1982, 2011). An identification task requires that the listener be able to detect and discriminate the stimulus. This means that the listener is able to distinguish the sound from all other sounds, for example: same or different, softer or louder, higher or lower. Then the listener should be able to uniquely identify (i.e. name) the stimulus (Small, 1973). Identification is thus a higher level task than is detection and/or discrimination. The listener may indicate identification of a spoken item by repeating, pointing to a picture/object, or writing the word or sentence that was perceived

An identification response may be segmental, where the listener describes the spoken item (Erber, 2011). For example, the child points to a picture of bola (a ball) after hearing the spoken word 'bola'. Or, the identification response may be suprasegmental in that the child identifies the general class of the speech stimulus based on its suprasegmental characteristics such as the rhythmic pattern of the speech stimulus. The listener describes the word's acoustic/rhythmic pattern without actually naming it. For examples, the child points to a symbolic representation of a word with a 4-beat rhythmic pattern (a waveform with four short pulses) after hearing the spoken word, such as *kupu-kupu* (butterfly), *topi-topi* (hats) and *kuda-kuda* (horses). Or, the child points to a symbolic representation of a word with a 2-beat rhythmic pattern (two short pulses) after hearing the spoken word, for example, *gigi* (teeth) and *sapi* (cow).

The identification subtest of INDO-SPASP will be used to distinguish between those children who can perceive the spectral/segmental components of words and those who can perceive only suprasegmental components. Many English speech perception tests for profoundly deaf children that were created for this purpose were subdivided into stress pattern categories. This is because English is a lexical stress language. Stress patterns are known to English speaking listeners and stress is used in auditory word recognition. Examples of this type of test are Monosyllable, Trochee, Spondee (MTS) (Erber & Witt, 1977); Auditory Numbers Test (ANT) (Erber, 1980) and Early Speech Perception (ESP) (Moog & Geers, 1990). In the Indonesian language, however, stress is communicatively irrelevant and it is not used in auditory word recognition.

Stress pattern is part of the suprasegmental/prosody component of speech. Suprasegmental patterns in English and Indonesian are discussed below.

5.2.1 Suprasegmental (Prosody)

When analysing speech it is usual to talk about segmental and suprasegmental (prosodic) information. The segmental information is related to the spectral features of vowels and of consonants (phonemes) and the suprasegmental information is related to the time-intensity pattern of syllables, words, phrases and sentences.

The main components of prosody in English are fundamental frequency, duration and intensity (Lieberman, 1966; Crystal, 1969; Lehiste, 1970). However, the terms used may differ, depending upon whether they are being considered from the speaker's point of view (physiological/production), the listener's point of view (perception), or as an acoustic manifestation (measurement). Table 5.3 was adapted from Lehiste (1970) to demonstrate this difference in terminology.

Table 5.3 Terminology describing suprasegmental feature according to the reference point of view and linguistic function

	Reference Point of View			**Linguistic Function**	
Suprasegmental	Speakers/ Physiological/ Production	Acoustic Manifestation/ Measurement	Listener's/ Perception	Word Level	Sentence Level
Quantity Features	Timing of articulatory sequences	Time dimension of the acoustic signal	Perception of duration	Quantity	Tempo

	Reference Point of View			Linguistic Function	
Tonal Features	Phonation	Fundamental Frequency	Perception of Pitch	Tone	Intonation
Stress Features	Effort level of manipulation of physiological mechanism	Intensity and amplitude	Perception of loudness and stress	Word stress	Sentence-level stress

The term that will be used in this study is the term from the listener's point of view or the *perception* point of view and it is only perception of stress that will be discussed further, as it is the relevant part of this study.

5.2.2 Perception of stress

In a Lexical stress language such as English and Dutch, stress patterns are known to listeners and stress is used in auditory word recognition. In English, words like FOREbear (means: ancestor; noun) and forBEAR (means: to cease or refrain from; verb) or TRUSty (means: faithful, trustworthy; adjective) and trusTEE (means: custodian, fiduciary; noun) (stressed syllables are capitalised) are mutually distinguishable by their stress patterns. (These examples were taken from Van Zanten and Van Heuven, 1998.) In many western languages, stress information is important in speech perception. In Indonesian, as opposed to many Western languages, stress is not distinctive: there are no words containing the same sequence of vowels and consonants that differ in their patterns and consequently in their meanings. Van Zanten and Van Heuven (1998) found that word stress was not used by Indonesian listeners to differentiate between words. Therefore, they concluded that stress is communicatively irrelevant and essentially free in Indonesian Language. As mentioned above, Dardjowidjojo (1978), also stated that word stress is not significant phonemically. Improper stress does not alter the meaning; it only makes the word sound unusual or unfamiliar. The amount of time required to say a sentence is determined by the number of the syllables within the sentence. The difference in duration between more prominent and less prominent syllables is comparatively small in Indonesian (Van Zanten & Van Heuven, 1997). This is different from English, in which the overall timing is heavily influenced by the number of stressed syllables in the sentence (Darjowidjojo, 1978). Indonesians are relatively tolerant as regards stress and its position (Van Zanten & Van Heuven, 2004). Neither the form nor the position of the accent-lending pitch movement which in stressed languages is associated with the stressed syllable seems to be of crucial importance to Indonesian listeners. Indonesians, however, can distinguish prosody cues but do not use them to identify (parts of) words. Stress does not play role in auditory recognition of words in Indonesian people who speak Indonesian language. Very often the Indonesian language is pronounced on the basis of regional dialects (languages) which may have different prominence patterns, but these prominence patterns do not influence the meaning. As a consequence, Indonesians when speaking Indonesian language do not pay attention to prominence location (Lapoliwa, 1981; Van Heuven & Van Zanten, 1997). However, the situation may be different for, for instance, Indonesian people who speak the Toba Batak language, and who do use stress in their regional language (Roosman, 2004).

In the view of most linguists in the past, the primary (or main) stress in the base words of Indonesian language was on the penultimate syllable (Dardjowidjojo 1978; Teeuw 1978). Van Zanten and Van Heuven's study (2004) also found a preference in the Indonesian language for primary prominence on phonologically heavy pre-final syllables. However, their study indicated that sets of target words which included stimuli with heavy pre-final syllables were not identified any better than other sets. It was therefore concluded that although Indonesian listeners may prefer prominent (heavy) pre-final syllables, such a preference has no relevance in speech communication. Further study by Van Zanten and Goedemans (2009) confirmed that there is no word-based stress in Indonesian.

5.2.3 Perception of syllables by profoundly hearing-impaired children

One of the acoustic features of speech that is available to the profoundly hearing impaired through the time intensity pattern is the number of syllables in a word, phrase, or sentence. By using syllable pulses (mainly the strong acoustic energy pulse in vowels) they would be able to discriminate many from few syllables (Zeisser & Erber, 1977).

Counting syllables may seem to be a simple task to perform, requiring only the ability to count or to discriminate between one or more beats. However, because the principal information available for their speech processing are only the pulses (i.e. the gross intensity envelope variations) in the acoustic pattern (also called time-intensity or vibratory or waveform/speech envelope pattern), perception of the syllabic patterns in words by profoundly hearing-impaired children differs considerably from normal auditory perception.

Several studies have indicated that profoundly hearing-impaired children do not actually hear the sounds presented to them during audiometric testing, but instead perceive the stimulus through vibrotactile receptors in their outer and middle ears (Bekesy, 1960; Nober, 1967; Erber, 1974; Risberg, 1977 in Erber, 1979). A study was conducted with normal hearing adults and profoundly hearing impaired children, to examine their abilities to count the number of syllables both vibrotactually (feel) and acoustically (listen). The performance of the profoundly hearing impaired children in the 'listen' condition was very similar to their performance in the 'feel' condition and was also similar to the performance of normal hearing adults in the 'feel' condition (Zeisser & Erber, 1977). The syllabic pattern or the number syllables perceived in these three instances were consistent with the waveform envelope shown on a storage oscilloscope screen after 200-600 Hz band-pass filtering (to simulate the sensitivity of profoundly impaired ears).

However, all of the above three instances were very different from the 'listen' condition of normal hearing adults (normal auditory perception). One might expect that one-syllable words exhibit a single energy pulse, two-syllable words exhibit two distinct pulses, three-syllable words exhibit three distinct pulses, and so on. However, variation within syllables such as syllable boundaries and unstressed vowels of some words may cause an energy pulse to be obscured, so that the syllabic pattern or the number of syllables perceived by profoundly hearing impaired children is different from the number of syllables as commonly defined by a dictionary or as heard by normal hearing listeners. For example, *lemon*, *running* and *hammer* are bisyllabic words as defined by a dictionary. They may appear, however, as monosyllabic patterns when they are perceived by profoundly hearing impaired children (Zeisser & Erber, 1977; Erber, 1979). This phenomenon tends to occur whenever syllables are joined by continuant consonants with low frequency energy, such as /l,r,m,n,w,j/.

In summary, perception of the syllabic patterns in words by profoundly hearing impaired children differs considerably from normal auditory perception. Therefore, in creating this type of test, one cannot develop a list of words based merely on the dictionary criteria of syllabification. One needs also to take into account the number of perceived syllable pulses in the waveform envelope.

Two examples of similar existing tests are: (i) MTS (Monsyllabic, Trochee, Spondee) (Erber & Witt, 1977); and (ii) the MED-EL's EARS (Evaluation of Auditory Responses to Speech) test battery that adopted MTS in the Malay language (Mukari & Abdul, 2008). These tests are discussed below.

5.2.3.1 Monsyllabic, Trochee, Spondee (MTS) (Erber & Witt, 1977)

The purpose of this test is to distinguish between those children who can perceive the spectral/segmental components and those who can perceive only time intensity patterns/suprasegmental components of speech. MTS consists of 30 stimulus words, 10 monosyllables, 10 trochees and 10 spondees. The bisyllabic words of each trochee and spondee, that were selected, were separated by a stop consonant. Thus, all of the bisyllabic words in MTS have a waveform with two distinct pulses and are likely to be perceived as two syllables vibrotactually or by the profoundly hearing impaired listeners, but with different stress patterns (strong-weak; strong-strong). Below is the list of 30 stimulus words.

Monosyllables	**Trochees**	**Spondees**
Broom	Button	Airplane
Chair	Chicken	Bathtub
Cup	Letter	Doghouse
Foot	Mountain	Firetruck
Girl	Paper	Goldfish

Monosyllables	Trochees	Spondees
Horse	Pocket	Icecream
Leaf	Rabbit	Lunchbox
Man	Spider	Popcorn
Pig	Turtle	Raincoat
Sock	Water	Toothpaste

Figure 5.3 MTS (Erber & Witt, 1977)

Each stimulus word was recorded on a separate Bell and Howell Language Master audio tape-card. Both syllable duration and separation in bisyllabic words were carefully monitored on a storage oscilloscope to maintain similarity within each word-pattern category. The duration of stressed syllables in all of the stimulus words ranged between 250 and 450 msec, while the duration of unstressed syllable peaks was 250-350 msec in the trochees and 550-650 msec in the spondee words. Care was taken to include a variety of vowels and consonants, although perfect phonetic balance was not attempted.

5.2.3.2 MED-EL Evaluation of Auditory Responses to Speech (EARS) Malay adaptation test (Mukari & Abdul, 2008)

The EARS Malay adaptation speech test was produced by the University Kebangsaan Malaysia (UKM) & MED-EL Innsbruck, Austria. One of the subtests is Monosyllabic, Trochee, Pollysyllabic (MTP), which is an adaptation of the MTS test (Erber & Witt, 1977). MTP consists of words with different number of syllables and it has three forms:. (i) MTP 12 consists of 12 words; (ii) MTP 6 consists of six words; and (iii) MTP 3 consists of three words in combination.

The methodology in developing the MTP test was unclear. It is noted, however, that the construction of the word list is not appropriate in terms of the suprasegmental characteristics of the words in MTP 6 and MTP 12.

For example, in MTP 6, the word classification for the suprasegmental characteristics of the words in the second and third columns is not appropriate. In the second column, the words *bola* and *kasut* are both bisyllabic words, but their suprasegmental patterns are very different. The rhythmic/suprasegmental pattern of ‘Bola’ is more like a long single pulse because the first syllable and the second syllable is joined by /l/ which is a continuant consonant with low frequency energy. In contrast, the rhythmic pattern of the word *kasut*, would be two pulses because the first syllable and the second syllable are joined by the voiceless fricative /s/. The fricative/s/ creates a gap between the two syllables and therefore, kasut has two separate waveforms/pulses. (Figure 5.5 is a simple graphical representation of the pulses/beats of the words.)

In the third column, the words *Kereta* and *Harimau* are both trysyllabic. However, the rhythmic/suprasegmental pattern of *Kereta* as heard by the profoundly hearing impaired would be a double pulse which consists of one single long pulse (the first syllable and the second syllable are joined by a continuant consonant with low frequency energy /r/) and a short single pulse (the second syllable and the third syllable are joined by a voiceless stop consonant /t/). The rhythmic pattern of the word ‘harimau’, would be a very long single pulse (the first syllable and the second syllable are joined by /r/ and the second syllable and the third syllable are also joined by a continuant consonant with low frequency energy /m/) (Figure 5.5).

	Jam	**Beg**	**Bola**	**Kasut**	**Kereta**	**Harimau**
Jam						
Beg						
Bola						
Kasut						
Kereta						
Harimau						

Figure 5.4 MTP 6

Word	Waveform	Rhythmic Pattern (Pulses)
Jam		Single pulse
Beg		Single pulse
Bola		Single pulse
Kasut		Two pulses
Kereta		Two pulse (one long pulse and one short pulse)
Harimau		Single pulse (a very long single pulse)

(Words were recorded by the present author using an adult female native speaker of Malay who was studying in Sydney and had standard Malay pronunciation).

Figure 5.5 Simple graphical representations of pulses of words in MTP 6

Clearly, the MTS test was adapted to the Malay language (Bahasa Melayu) based only on the dictionary criterion of syllables without understanding that the original English MTS test was developed based on both syllabic (spectral cues) and suprasegmental perceived cues. Thus, the current Malay version cannot distinguish accurately between those children who can perceive spectral/segmental components and those who can perceive only suprasegmental components.

Unless, MTP has a different purpose than the original MTS, it is not valid to use it for assessing the suprasegmental perception ability of children with profound hearing loss.

The identification subtest of INDO-SPASP developed in this study considered: (i) the number of syllables based on dictionary criteria of syllabification (i.e. requires spectral-temporal/segmental cues); and (ii) the number of perceived (either auditorily or vibrotactually) syllable pulses in the waveform envelope (suprasegmental) by profoundly hearing impaired children. The identification subtest of INDO-SPASP consists of a list of bisyllabic words with a single energy pulse, a list of bisyllabic words with two distinct pulses and a list of polysyllabic words with four distinct energy pulses. Unlike, many English speech perception tests mentioned above, the word lists in this subtest are not divided into stress pattern category because, as mentioned earlier, stress in Indonesian language is not used in auditory word recognition and it is communicatively irrelevant.

5.3 Research question

This study aimed to answer the following research question:

> **Can the identification subtest of INDO-SPASP distinguish between those children who can perceive spectral/segmental components and those who can perceive only suprasegmental components?**

CHAPTER 6:
INDO-SPASP METHODOLOGY

6.1 Developing the stimulus material

6.1.1 Instrumentation

***Praat*^{TM} Software V5.0.14 (Boersma & Weenink, 2008).** *Praat*^{TM} (also the Dutch word for 'talk') is a free scientific software program that offers a wide range of procedures, including speech analysis, articulatory synthesis, and neural networks. It has been designed and continuously developed by Paul Boersma and David Weenink, Institute of Phonetic Sciences, University of Amsterdam.

HELOS (hardware and software) (Erber 2008). The Hearing Loss Simulator (HELOS) system is based on a simple model of hearing loss. It simulates the perceptual effects of hearing loss, and produces similar auditory effects. The hardware version was used in this study because the software version does not simulate profound hearing loss. The HELOS hardware version is an analogue device which is intended to approximate the effect of damage to the auditory system (Figures 6.1 & 6.2).

Figure 6.1 HELOS hardware (reproduced with permission from N. P. Erber)

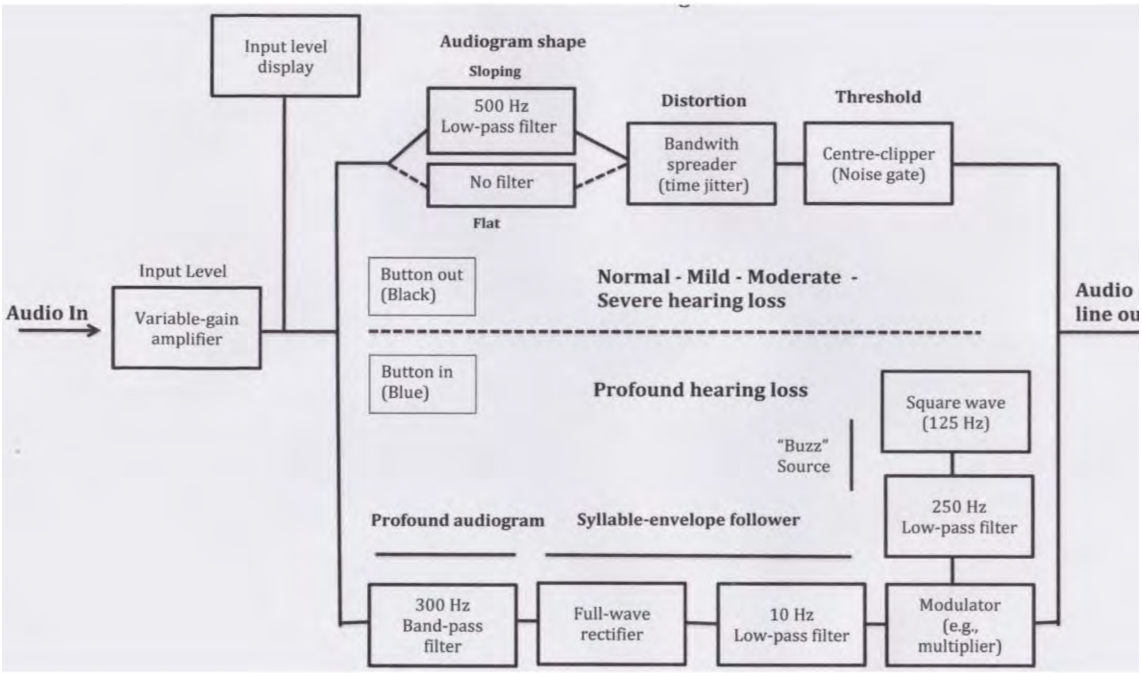


Figure 6.2 HELOS hardware diagram

6.1.2 Procedure

Twenty-two concrete bisyllabic words and 10 concrete polysyllabic words were selected from the corpus of 266 concrete Indonesian words. Concrete words are words that are representable by pictures. The words were familiar to severe-profoundly hearing impaired children as young as 7.6 years old (Section 4.2.2). Their acoustic similarities and differences were assessed with *Praat*^{TM} *V.5.0.14* and by listening to the words through HELOS.

Each word was initially examined by observing the continuity of its waveforms in *Praat*^{TM} (Appendix G). A single continuous waveform of speech (uninterrupted by gaps), in this study, is defined as *a single syllable pulse*. For example, *Bola* (ball) has one pulse because it produces a single continuous waveform. Other examples are *Topi* (hat) and *Buku-Buku* (books). *Topi* has two pulses. The first pulse is the waveform before the silent gap and second pulse is the waveform after the gap. *Buku-Buku* (books) has four pulses. The waveform shown in the graphical representation (Figure 6.3) is a time compressed waveform so that the beats of the syllable pulses are emphasised.

The polysyllabic word list was created based on the bisyllabic words with two pulses. In Indonesian, a repeated word such as *Buku-Buku* indicates the plural form of buku (books). These syllable pulses are different from glottal pulses. A *Glottal pulse* is an individual opening and closing of the vocal folds which might happen 200 times per second for an adult female speaker (F0).

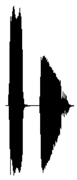
Word	Waveform	Pulses
Bola		Single pulse
Topi		Two waveforms interrupted by a gap/ two pulses
Buku-Buku		Four pulses

Figure 6.3 Graphical representation of pulses of words in INDO-SPASP

Three word lists were created that differed in syllable patterns. Each list contained words that differed spectrally as follow:

1. Eleven bisyllabic words with one long beat (a waveform with one long single pulse). This group consists of bisyllabic words with a medial sonorant continuant consonant which include /l/, /η/, /m/, /n/. The words include: (1) *bola* (ball); (2) *bulan* (moon); (3) *bunga* (flower); (4) *gelas* (glass); (5) *rumah* (house); (6) *semut* (ant); (7) *tomat* (tomato); (8) *mulut* (mouth); (9) *nanas* (pineapple); (10) *nenek* (grand mother); and (11) *ular* (snake).

2. Eleven bisyllabic words with two short beats (a waveform with two short single pulses). This group consists of bisyllabic words with a medial stop consonant such as /g/, /k/, /p/, /d/ and /t/ or voiceless fricative such as /s/. The words include: (1) *gigi* (teeth); (2) *kaki* (teeth); (3) *kursi* (chair); (4) *nasi* (rice); (5) *sapi* (cow); (6) *topi* (hat); (7) *sapu* (broom); (8) *susu* (milk); (9) *buku* (book); (10) *kuda* (horse); and (11) *pintu* (door).

3. Ten concrete polysyllabic words with four short beats (a waveform with four short single pulses) were selected. This group consists of mostly repeated bisyllabic words with medial stop consonants which include /p/, /k/, and /d/ or voiceless fricatives such as /s/. The words include: (1) *kupu-kupu* (butterfly); (2) *buku-buku* (books); (3) *kaki-kaki* (legs); (4) *topi-topi* (hats); (5) *sapi-sapi* (cows); (6) *kapal-kapal* (boats); (7) *sikat-gigi* (tooth brush); (8) *meja-meja* (tables); (9) *kursi-kursi* (chairs); and (10) *kuda-kuda* (horses). Most repeated bisyllabic words in this list indicate plurality except *kupu-kupu*. *Sikat-gigi* is the only non-repeated word and it does not indicate plurality.

Polysyllabic words with four short pulses were selected because other polysyllabic words such as concrete words with three pulses are uncommon in the Indonesian language. Monosyllabic words are not included, because most of words in the Indonesian language are bisyllabic or polysyllabic and most of monosyllables in Indonesian are nonsense syllables or exclamations.

The three word lists were then recorded in the Macquarie University Linguistics Studio. The speaker was the author (female's voice). The speech was recorded using an AKG C535EB condenser microphone connected to a Mackie 1402 VLZ mixer, produced by LOUD Technologies, Inc. Microphone placement was 150 mm in front and at 45 degrees to the side of the speakers mouth, to avoid turbulent airflow. Signals were captured through the M-Audio Delta 64 PCI digital recording interface attached to *MS Windows*™ PC. Recording software was 'Cooledit'. Recordings were made at a 44.1 kHz sample rate and 16 bit quantisation. The sound files were copied to Data CD. *Praat*™ software was used to segment the sound files into individual tokens.

Then, the words in each list were put through HELOS to confirm the selection by listening to the rhythmic pattern of each word approximately as heard by a profoundly hearing impaired person. It is realised that the actual perception of the test items by profoundly hearing impaired person may differ from the sounds processed by HELOS. However, it was felt that the equipment would produce useful information with regard to the design of test materials.

Speech signals were passed through HELOS and the output speech signals were recorded. The signal levels were restored for peak levels similar to the input signal. Due to the processing of the speech signals, it was not possible to restore Leq levels. The processed signals were fed via the Mackie mixer into the same hardware combination listed above.

The HELOS output showed that the rhythmic patterns of the word lists were consistent with the waveform pulse patterns observed in *Praat*™ software. Words in the 1st list sounded as single pulses, words in the second list sounded as two pulses and words in the third list sounded as four pulses (Appendix H). Five words within each list with the most similar sounding rhythmic pattern were selected. Care was taken to include a variety of vowels and consonants, although perfect phonetic balance was not attempted.

- List 1 consists of five words with a long single pulse: (1) *rumah* (house); (2) *bola* (ball); (3) *nanas* (pineapple); (4) *gelas* (glass); and (5) *nenek* (grandmother).

- List 2 consists of five words with two short pulses: (1) *gigi* (teeth); (2) *pintu* (door); (3) *sapi* (cow); (4) *topi* (hat); and (5) *kursi* (chair).

- List 3 consists of five words with four short pulses: (1) *buku-buku* (books); (2) *kapal-kapal* (ships); (3) *sapu-sapu* (brooms); (4) *topi-topi* (hats); and (5) *kuda-kuda* (horses).

A stimulus/response matrix was developed (Figure 6.4) and a picture card was created for each word. Pictures were taken from the collection of INDO-SPRITT's pictures (Figures 6.5a, 6.5b & 6.5c).

		Stimulus														
		Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Responses	Rumah															
	Bola															
	Nanas															
	Gelas															
	Nenek															
	Gigi															
	Pintu															
	Sapi															
	Topi															
	Kursi															
	Buku-Buku															
	Kapal-Kapal															
	Sapu-Sapu															
	Topi-Topi															
	Kuda-Kuda															

Figure 6.4 Stimulus/response matrix test form sheet

Figure 6.5a Pictures of words with a long single pulse/beat

Figure 6.5b Pictures of words with two short pulses/beats

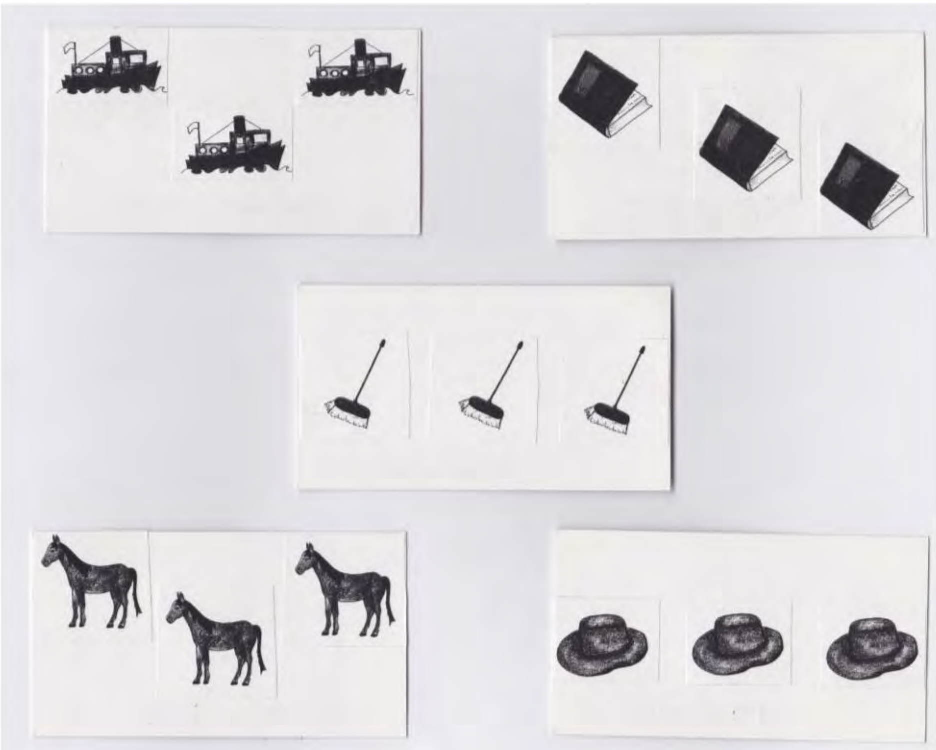


Figure 6.5c Pictures of words with four short pulses/beats

Below is a simple graphical representation of the pulses/beats of the words to show an impression of the kind of pulses that would be perceived by a profoundly hearing impaired person (words after they had been processed through HELOS) (Figures 6.6a, 6.6b & 6.6c).

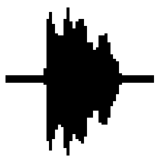
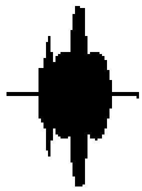
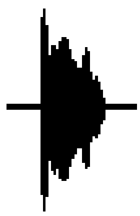
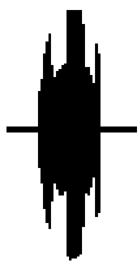
No	Words	Normal	HELOS
1	Rumah		
2	Bola		
3	Nanas		
4	Gelas		
5	Nenek		

Figure 6.6a Bisyllabic waveforms – single pulse

The HELOS waveform is not intended to be a full representation of speech perception for the profoundly hearing impaired child. It does not show the exact extent of quantisation (number of levels of intensity that can be discriminated) by a hearing impaired child. HELOS is meant to give an impression of the effect of the hearing impairment to the parents of the child.

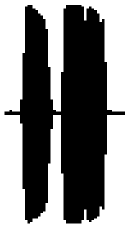
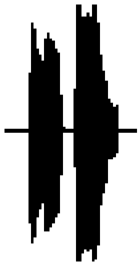
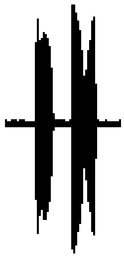
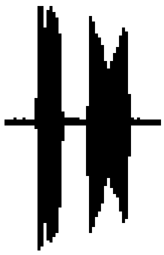
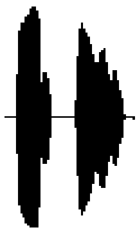
No	Words	Normal	HELOS
1	Gigi		
2	Pintu		
3	Sapi		
4	Topi		
5	Kursi		

Figure 6.6b Bisyllabic waveform – two pulses

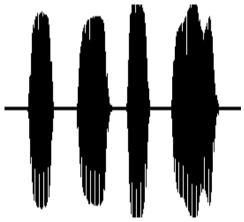
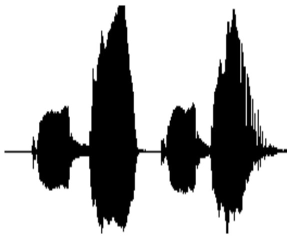
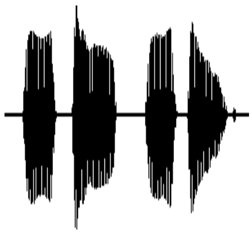
No	Word	Normal	HELOS
1	Buku-buku		
2	Kapal-Kapal		
3	Sapu-Sapu		
4	Topi-Topi		
5	Kuda-Kuda		

Figure 6.6c Polysyllabic waveforms – four pulses

6.2 Pilot Study

6.2.1 Subjects

A pilot study performed at the Karnamanohara Deaf School in Yogyakarta, Indonesia was undertaken with seven Indonesian children with severe to profound hearing loss in the age range between eight to 13 years. This group included two severely hearing-impaired children with mean hearing threshold levels for the better ear 83 and 86 dBHL, and five profoundly hearing-impaired children with mean hearing threshold levels for the better ear between 91 and 115 dBHL.

The children who participated in this pilot study were the ones who wore hearing aids more regularly compared to other children. Five children had binaural hearing aids and the other two each children had a monaural hearing aid.

At the time of the study, 16 children had hearing aids out of 80 students studying in that school. Families of most children in that school could not afford to buy hearing aids or batteries.

The cause of the hearing loss was unclear (whether it was acquired or congenital). The school could not provide the history of the children's hearing loss. However, many of the students came to the school before the age of five. For those who came after the age of 5, their speech was unintelligible. Therefore, it is likely in most cases that their hearing loss occurred prior to language acquisition. The communication method taught in the school was auditory-oral (auditory-visual). However, given that the majority of the students did not have hearing aids, lip-reading, gesture and body language became the primary communication method used by the children.

6.2.2 Testing procedure

The condition of the testing room was less then optimal. The partition between the rooms in that school was made from a thin layer of webbing bamboo (Appendix I). Therefore, noise from the other rooms was often audible and intrusive. Each test had to be stopped several times, and restarted when the noise stopped.

Each child was tested individually. Hearing aids were checked to ensure they were working properly. A speech detection test (Ling Sound Test) was performed. The phonemes that were tested include /a/, /i/, /u/, /m/, /s/ and /ʃ/. All of the words and pictures were ensured to be familiar to all of the children. The subjects were then given practice listening to the stimulus words and auditory-visual instructions were given to the children. Each child then listened to the words and pointed to corresponding pictures. Finally, the 15 words were presented randomly with live voice to the subjects. Live voice presentation was monitored with a sound level meter at about 65-70 dBA.

All responses were recorded by the examiner on confusion matrices which displayed both correct identifications and errors.

CHAPTER 7:
INDO-SPASP RESULTS AND DISCUSSION

7.1 Results and discussion of phoneme detection

Of the seven children, only one child (subject NK) could detect all of the phonemes (Table 7.1). This indicates that NK's hearing aid provided enough gain and output level to enable her to detect acoustic energy, at least up to 5000 Hz, as the dominant spectral energy of voiceless fricative/s/ in adult females is around 5000 Hz (the speaker was the author). Subject SR could detect most of the six phonemes, except phoneme/s/. This result indicates that SR's hearing aid enabled him to detect acoustic energy up to 3000 Hz, because the dominant spectral energy for a voiceless fricative /ʃ/ is around 3000 Hz.

Table 7.1 Phoneme detection

No	Age	Subjects	3 FAHL		Phonemes					
			Right	Left	/a/	/i/	/u/	/m/	/s/	/ʃ/
1	11	NK	83.3 dBHL	80 dBHL	✓	✓	✓	✓	✓	✓
2	11	SR	95 dBHL	91 dBHL	✓	✓	✓	✓	×	✓
3	10,9	YK	93 dBHL	86 dBHL	✓	✓	✓	✓	×	×
4	13	RA	105 dBHL	98 dBHL	✓	✓	✓	✓	×	×
5	8	T	113 dBHL	103 dBHL	✓	✓	✓	×	×	×
6	11	GHM	108 dBHL	108 dBHL	✓	✓	✓	✓	×	×
7	9,8	LP	No response	115 dBHL	✓	✓	✓	×	×	×

Note: ☑: able to detect; ☒: unable to detect

The other three children (RA, GHM and YK) could not detect the voiceless fricative consonant sounds but they all could detect the vowel sounds and nasal /m/sound. These results indicate that the children's hearing aids enable them to detect acoustic energy within at least up to the range of first formant (F1). However, these children may not be able to distinguish between phoneme /i/ and /u/ because they could not hear higher frequencies and therefore higher vowel spectral (formant) peaks. To distinguish between these two phonemes (identify the two phonemes) the child must be able to hear the second formant for /i/ which has acoustic energy around 2400 Hz (Figure 5.1 & Table 5.2).

The other two children (T and LP) could detect all of the vowel sounds but could not detect the nasal lower frequency murmur of the /m/ sound. According to Ling (2002), an inability to hear /m/ at a distance of approximately three metres indicates that a child is unlikely to adequately hear other low-frequency sounds that are important for the development of normal speech, because the peak energy of /m/is at 250-300 Hz (Cox, 2008) and F1 of Indonesian phonemes /i/ and /u/ around 200-300 Hz (Zanten, 1989). However, these children could detect all of the corner vowel sounds (/a/, /i/ and /u/). The possible explanation of this phenomenon is that it could be due to the fact that the nasal energy at very low frequencies is either weakly audible or not audible for these children. Nasals are lower in intensity than vowels due to damping (Cox, 2008).

Despite the fact that some of the children could not hear consonant sounds, these results indicated that it was appropriate to proceed with Subtest 2 to all children because they can all detect the vowel sounds. One of the purposes of the detection test is to determine if it is reasonable to proceed to Subtest 2. This is to avoid the situation where the person administering the test is attempting to test a child’s speech perception abilities with inaudible sound. Although, some children in the pilot study could only detect vowel sounds, it was still appropriate to proceed to Subtest 2 because the words consist of different vowels. The ability to detect different vowels may be used as an important cue to distinguish different words for those who can still utilise some spectral cues in word perception. For those whose ability to perceive speech or words is limited to time and intensity information, this detection test is useful to see if they can hear any sound at all (intensity of the sound). There is no detection effect for Indonesian vowel phonemes based on their duration, because Indonesian vowels have no significant long and short vowel distinctions as in English (Lapoliwa, 1981). The ability to perceive duration of speech (number of syllables) will be shown in Subtest 2.

Thus, it should be stressed that the subjects who are able to detect or to hear some sounds in the detection test, may not necessarily be able to utilise the spectral components to recognise or identify words. Similarly to the case of pure tone detection, the subjects are not necessarily able to utilise the spectral component of the pure tone to recognise speech. Detection is only the basic process of determining whether sound is present or absent (Small, 1973; Erber, 1982, 2011). But, to be able to identify words the listener should be able to detect and discriminate the stimulus. Being able to discriminate means that the listener is able to distinguish the sound from all other sounds, for example: same or different, softer or louder, higher or lower.

7.2 Results and discussion of word identification

The percent of words (segmental) and syllable patterns (suprasegmental) recognised correctly was plotted as a function of the 3 frequency average hearing threshold level (3 FAHL) of their better ear (Table 7.2 & Figure 7.15). The better ear was used because the speech perception testing was conducted live-voice and thus, it was assumed that the better ear would contribute more to the speech perception, although this may not always be the case with profound hearing loss (Cramer & Erber, 1974; Erber, 1974; Risberg et al., 1975; Erber, 1979, 1980; Plant, 1984).

Table 7.2 3 FAHL of the better ear and the score of suprasegmental and segmental performance.

No	Age	Subjects	Best 3 FAHL	Suprasegmental	Segmental
1	11	NK	80 dBHL	100%	86.7%
2	11	SR	91 dBHL	100%	86.7%
3	10,9	YK	93 dBHL	80%	20%
4	13	RA	98 dBHL	60%	40%
5	8	T	103 dBHL	46.7%	0%
6	11	GHM	108 dBHL	80%	46.7%
7	9,8	LP	115 dBHL	80%	26.7%

Most subjects wore two hearing aids except two subjects (YK and LP) who were only wearing a hearing aid in one ear and their 3 FAHL was used from the ear with hearing aid (Tables 7.2 & 7.3). The reason why YK had only one hearing aid was because the other hearing aid was broken and LP had no hearing in the right ear.

In general the identification scores obtained decreased markedly with increasing hearing loss (Table 7.2 & Figure 7.15). The children with hearing threshold levels better than 93 dB seem to have little difficulty identifying both the words (segmental) and the syllable patterns of the words (suprasegmental). The word identification score for Subject 1 (NK) and Subject 2 (SR)) with a hearing level equal or better than 91 dBHL was 86.7% and they performed this task with a high degree of proficiency, whereas the children with a hearing level worse than 91 dBHL appear to be using time and intensity information (syllable pattern) more than spectral cues. This result was consistent with the previous study by Zeisser and Erber (1977).

Table 7.3 3 FAHL of both ears and their hearing aids

No	Age	Subject	3 FAH:		Hearing aids	
			Right	Left	Right	Left
1	11	NK	83.3 dBHL	80 dBHL	Solo T Plus 411 DM	Solo T plus 411 DM
2	11	SR	95 dBHL	91 dBHL	Phonak Maxx 31 Forte	Phonak Maxx 31 Forte
3	10,9	YK	93 dBHL	86 dBHL	Solo T Plus 411 DM	No hearing aid
4	13	RA	105 dBHL	98 dBHL	Solo T Plus 411 DM	Solo T plus 411 DM
5	8	T	113 dBHL	103 dBHL	Phonak Extra 411 AZ	Phonak Extra 411 AZ
6	11	GHM	108 dBHL	108 dBHL	Solo T Plus 411 DM	Solo T plus 411 DM
7	9,8	LP	No response	115 dBHL	No hearing aid	Audio service Astral 23 HP

7.2.1 Audiograms and word list results

(1) Subject 1: NK (11 years old)

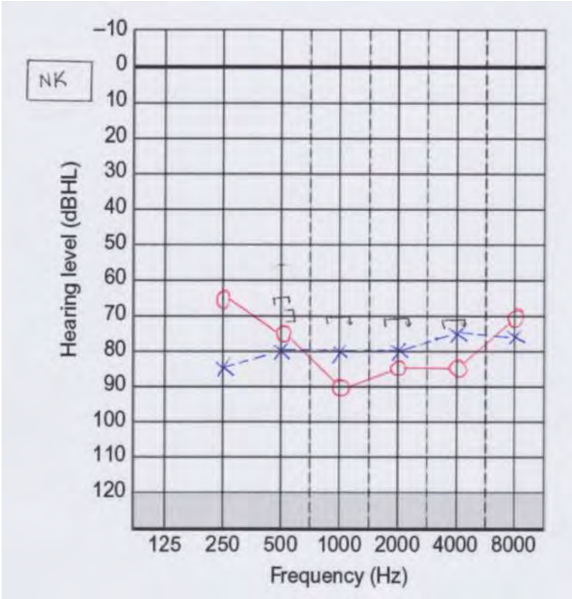


Figure 7.1 NK's audiogram

Test form sheet

NK (11 Y.O)

Responses \ Stimulus	Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah	•														
Bola		•													
Nanas															
Gelas				•											
Nenek			•		•										
Gigi						•									
Pintu							•								
Sapi								•							
Topi									•						
Kursi										•					
Buku-Buku											•				
Kapal-Kapal												•			
Sapu-Sapu													•	•	
Topi-Topi															
Kuda-Kuda															•

Figure 7.2 NK's word list results

Segmental score: 13/15x 100% = 86.7%; Suprasegmental score: 15/15 x 100% = 100%

(2) Subject 2: SR (11 years old)

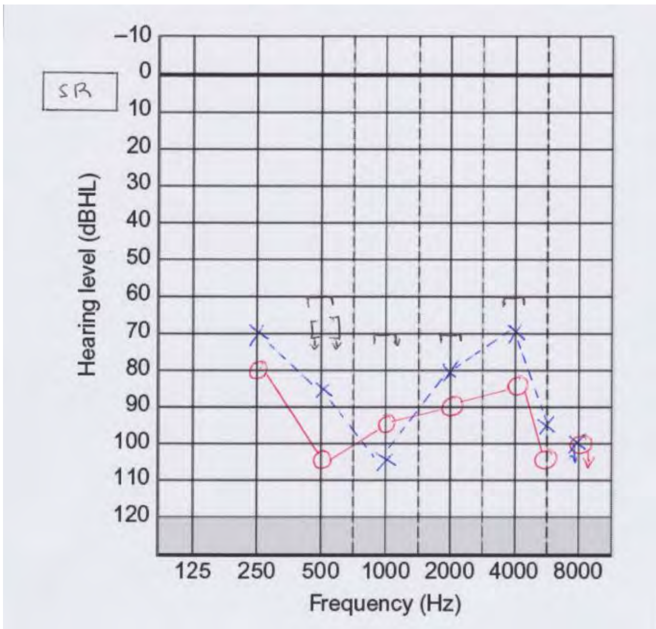


Figure 7.3 SR's audiogram

Test form sheet

SR (11,8 Y.O)

Responses \ Stimulus	Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah	•														
Bola		•		•											
Nanas															
Gelas															
Nenek			•		•										
Gigi						•									
Pintu							•								
Sapi								•							
Topi									•						
Kursi										•					
Buku-Buku											•				
Kapal-Kapal												•			
Sapu-Sapu													•		
Topi-Topi														•	
Kuda-Kuda															•

Figure 7.4 SR's word list results

Segmental score: 13/15 x 100% = 86.7%; Suprasegmental score: 15/15 x 100% = 100%

(3) Subject 3: YK (10 years and 9 month old)

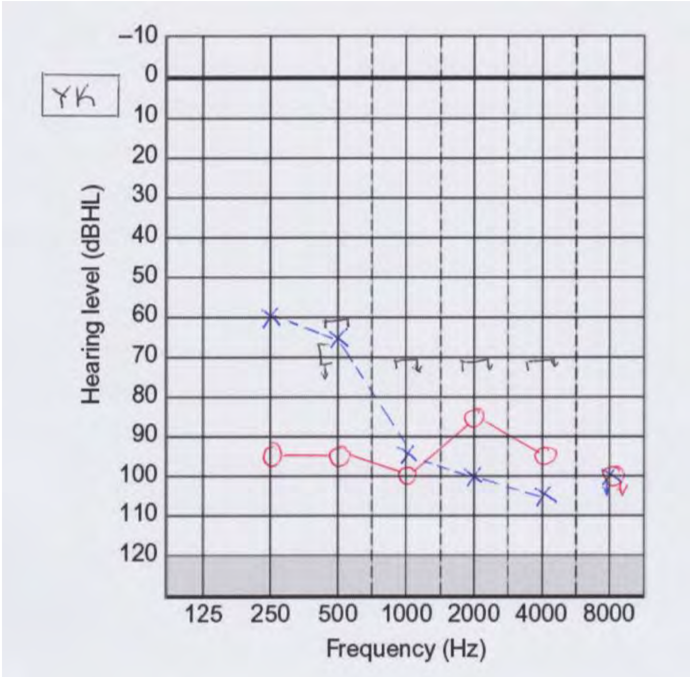


Figure 7.5 YK's audiogram

Test form sheet

YK (11 Y.O)

Responses \ Stimulus	Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah									•						
Bola	•			•											
Nanas			•		•										
Gelas		•													
Nenek						•	•								
Gigi															
Pintu															
Sapi															
Topi										•					
Kursi								•							
Buku-Buku															
Kapal-Kapal													•		
Sapu-Sapu											•				
Topi-Topi														•	
Kuda-Kuda												•			•

Figure 7.6 YK's word list results

Segmental score: 3/15 x 100% = 20%; Suprasegmental score: 12/15 x 100% = 80%

(4) **Subject 4: RA (13 years old)**

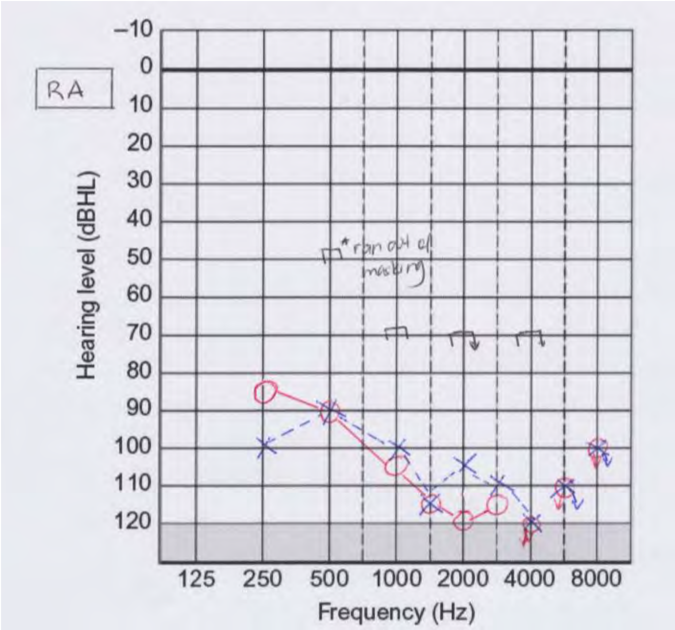
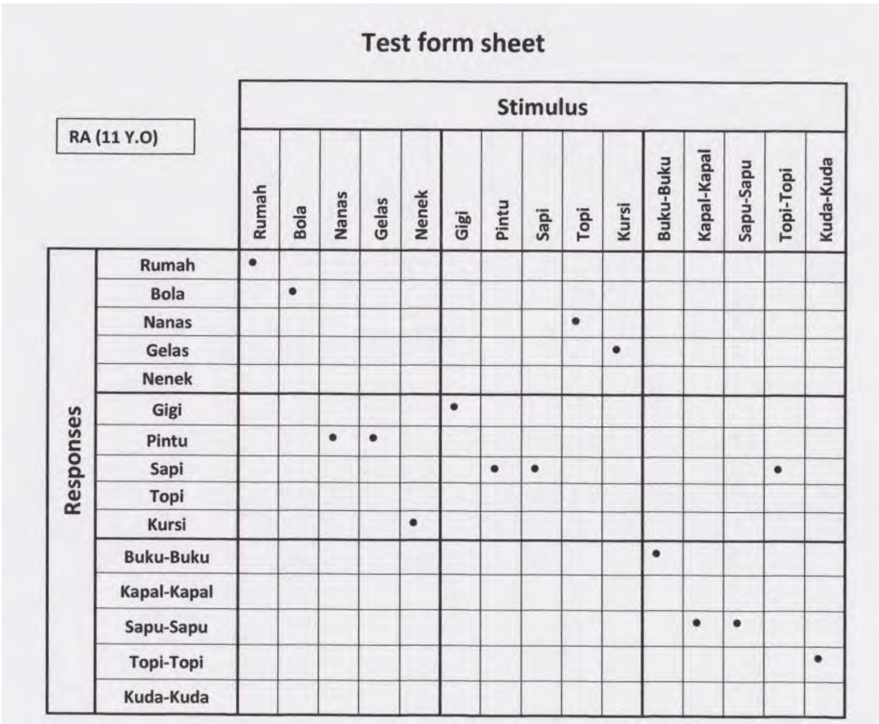


Figure 7.7 RA's audiogram



Test form sheet

RA (11 Y.O)

Responses \ Stimulus	Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah	•														
Bola		•													
Nanas									•						
Gelas										•					
Nenek															
Gigi						•									
Pintu			•	•											
Sapi							•	•						•	
Topi															
Kursi					•										
Buku-Buku											•				
Kapal-Kapal															
Sapu-Sapu												•	•		
Topi-Topi															•
Kuda-Kuda															

Figure 7.8 RA's word list results

Segmental score: 6/15 x 100% = 40%; Suprasegmental score: 9/15 x 100% = 60%

(5) Subject 5: T (8 years old)

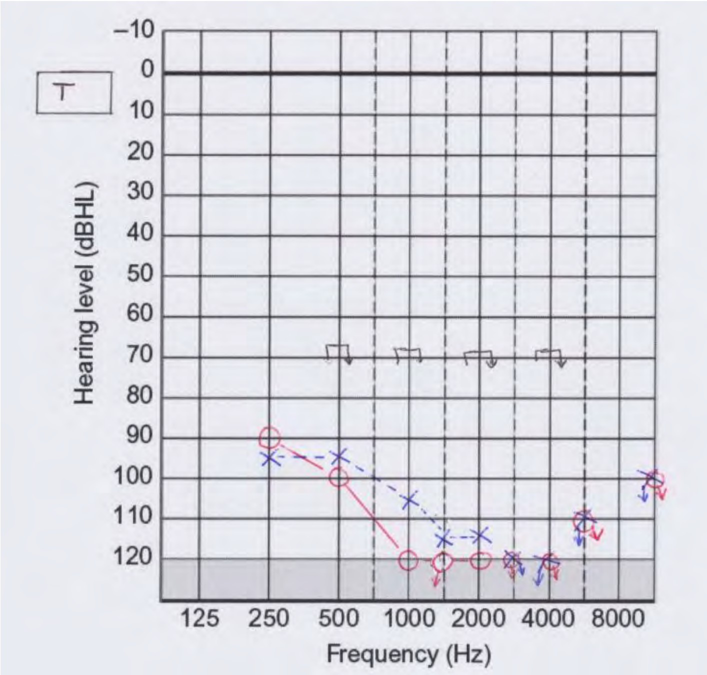


Figure 7.9 T's audiogram

Test form sheet

T (8 Y.O)

Responses	Stimulus: Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah															
Bola	•												•		
Nanas						•									
Gelas			•												
Nenek									•						
Gigi								•							
Pintu										•					
Sapi											•				
Topi															
Kursi		•													
Buku-Buku				•	•		•								
Kapal-Kapal															•
Sapu-Sapu												•			
Topi-Topi															
Kuda-Kuda														•	

Figure 7.10 T's word list results

Segmental score: 0/15 x 100% = 0%; Suprasegmental score: 7/15 x 100% = 46.7%

(6) Subject 6: GHM (11 years old)

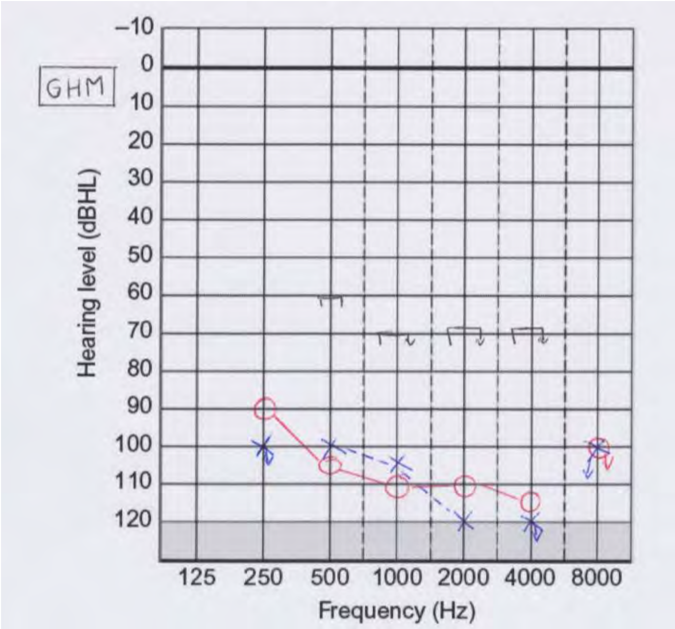
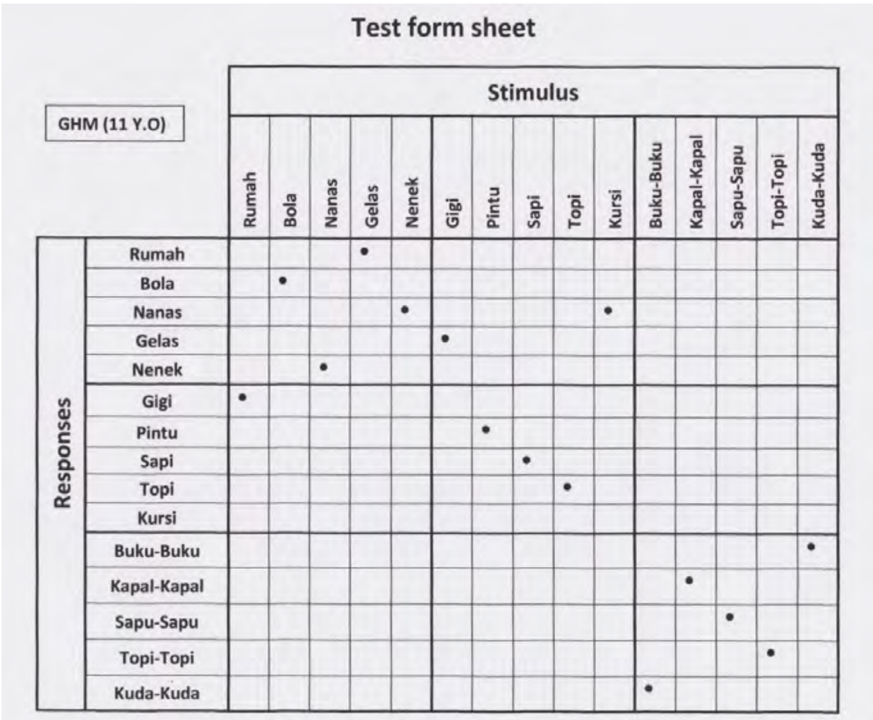


Figure 7.11 GHM's audiogram



Test form sheet

GHM (11 Y.O)

Responses	Stimulus: Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah				•											
Bola		•													
Nanas					•					•					
Gelas						•									
Nenek			•												
Gigi	•														
Pintu							•								
Sapi								•							
Topi									•						
Kursi															
Buku-Buku															•
Kapal-Kapal												•			
Sapu-Sapu													•		
Topi-Topi														•	
Kuda-Kuda											•				

Figure 7.12 GHM's word list results

Segmental score: 7/15 x 100% = 46.7%; Suprasegmental score: 12/15 x 100% = 80%

(7) Subject 7: LP (9 years and 8 months old)

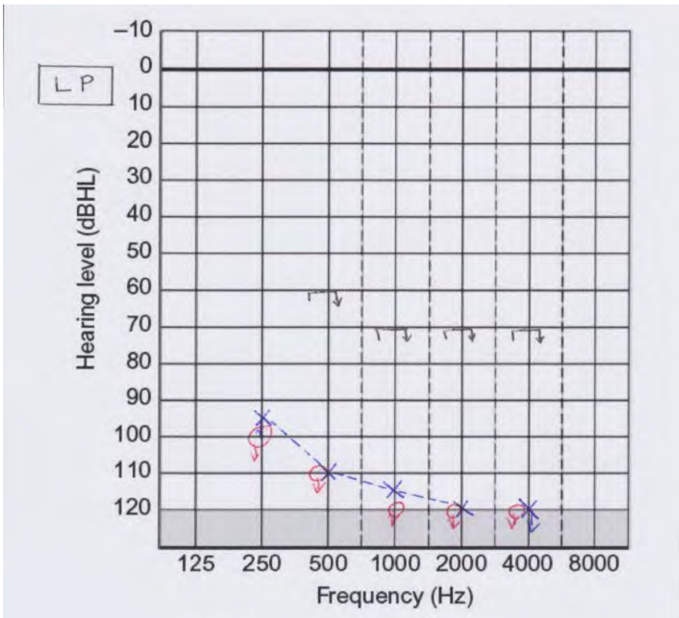


Figure 7.13 LP's Audiogram

Test form sheet

LP (9,8 Y.O)

Responses	Rumah	Bola	Nanas	Gelas	Nenek	Gigi	Pintu	Sapi	Topi	Kursi	Buku-Buku	Kapal-Kapal	Sapu-Sapu	Topi-Topi	Kuda-Kuda
Rumah		•													
Bola			•												
Nanas															
Gelas	•														
Nenek															
Gigi						•									
Pintu					•										
Sapi							•								
Topi								•	•						
Kursi				•											
Buku-Buku											•				•
Kapal-Kapal												•			
Sapu-Sapu															
Topi-Topi										•			•	•	
Kuda-Kuda															

Figure 7.14 LP's word list results

Segmental score: 4/15 x 100% = 26.7%; Suprasegmental score: 12/15 x 100% = 80%

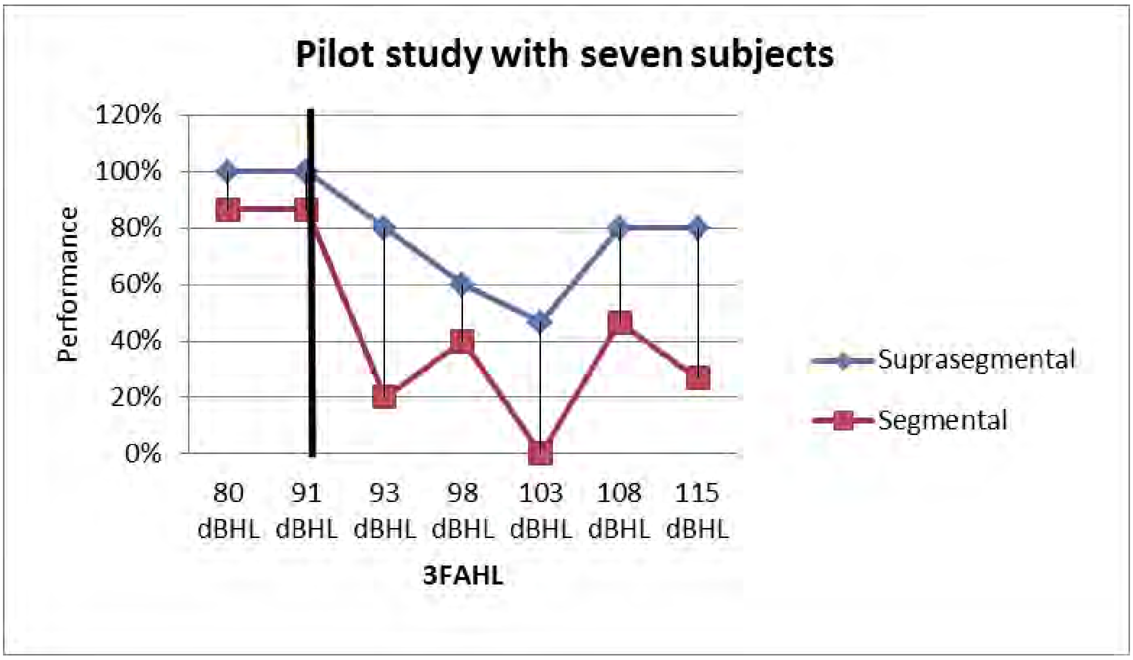


Figure 7.15 Percent of segmental and suprasegmental recognised correctly as a function of 3 FAHL of the better ear

The thick black line indicated the point at which the children with hearing loss over 91 dBHL identify words primarily on the basis of their syllable/suprasegmental pattern, whereas children with a less profound hearing loss can use segmental cues. The syllable pattern perception (suprasegmental) scores for children with hearing thresholds poorer than 91 dB, however, vary widely from 46.7% to perfect categorisation (100%). That is, children whose word recognition performance generally is poor (Figure 7.15) demonstrate large differences in word categorisation ability.

These differences in ability to perceive the syllable pattern could be due to impairment in some part of hearing pathway (peripheral and central) which might include hair cells damage, auditory neuropathy or auditory processing problem or differences in amount of practice or exposure to general listening. For example, Subject 5 (T) had the lowest score of all other subjects. Her segmental score was zero and the suprasegmental score was only 46.7% (Figure 7.10). This could be due to the impaired in some part of the hearing pathway but it was also likely that it could be due to the amount of practice and duration of listening/exposure to speech, since she was the youngest (8 years old) among the subjects.

As mentioned earlier, children with audiometric averages of 90 dB or greater demonstrated a range of speech perception abilities. An audiogram is not a good predictor of speech perception ability for profound hearing loss. For example, the audiograms of Subject 3 (YK) and Subject 2 (SR) were not much different (Table 7.2 and Figures 7.3 & 7.5) and yet, their perception abilities differed markedly. SR had a much better ability to use spectral cues, as his segmental score was 86.7% whereas YK's segmental score was only 20%. However, it is important to note that, apart from possibilities such as basic differences in function of the sensory mechanism (central and peripheral) or differences in the amount of practice or exposure to general listening, duration of loss and duration of using hearing aid, another factor that may also influence the result is the fact that SR had binaural hearing aid and YK had only a hearing aid in the right ear.

Another example of audiograms not being a good predictor of speech perception ability in this study is that LP's audiogram with 3 FAHL of 115 dBHL (Subject 7) was significantly worse than YK's audiogram with 3 FAHL of 93 dBHL (Subject 3), (and they both only had one hearing aid) but LP's ability to use spectral/segmental cues was similar or slightly better than YK. LP's segmental score was 26.7% and YK's segmental score was only 20%.

Furthermore, a child may demonstrate significantly different speech perception ability between the two ears, and yet the audiometric configurations of the two ears are very similar (Erber & Alencewicz, 1976). In this study the speech perception testing was conducted live voice and therefore, we did not have separate ear information. Thus, we did not know if both ears contribute equally in processing the words that the subjects listen to, or if one ear contributes more than the other, or if one ear had virtually no function.

It should be stressed, however, that performance on the test did not represent the child's optimal level of performance. It may be that the child's speech perception abilities can be greatly improved with appropriate training. Test results should be seen as base line data to be used in the planning of auditory training strategies and educational placement.

The issue of potentially high ambient noise levels in the test rooms, which might have affected the subjects speech scores, was addressed by monitoring the ambient noise levels with the sound level meter and halting the test procedure if the level exceeded 55 dBA.

Although many factors, such as parental support and availability of appropriate teaching skills may contribute to the placement of a child in a particular educational setting, as mentioned earlier, the criteria for the educational placement might take into account the child's segmental perception ability as follows:

- Segmental score ≥80% might indicate that the child is placed in an auditory/verbal class
- Segmental score between 50% to 80% might indicate that the child is placed in an auditory/oral class
- Segmental score 50% might indicate that the child is placed in a total communication class (Moore, 2014).

For example, if we use the results in this study as the basis of the placement, only two children (Subject 1 and Subject 2) would be placed in the auditory-verbal class (segmental score >80%) and the other five children would be placed in a total communication class because their segmental score is less than 50%.

The Karnnamanohara Deaf School (SLB-B Karnnamanohara) where the children were studying adopts aural-oral method. Sign language was discouraged. The students were placed based on their age. Preschool (1.8-3 years old), kindergarten (4-6 years old) and primary school (≥6 years old). The school used the national curriculum for hearing impaired children. For the primary school children, the mainstream curriculum was also integrated into the program.

It is hoped that the results presented in this study will show that the test will provide the hearing health care professional and teacher with important information on a child's auditory capacities and as such as should be a useful instrument in the overall management adopted with hearing-impaired children.

The results obtained by those children who appear to perceive only time intensity cues also indicate that they should derive benefit from hearing aid usage. The ability to detect syllable type, number and pattern offers useful information to supplement lip-reading and to assist in speech acquisition (Erber, 1971; Plant, 1984). This is particularly important in developing countries where a cochlear implant may not be an option because many of the hearing impaired children come from low to middle economic background and their health care system does not support the provision of cochlear implant. In developed countries where the health care system supports the provision of cochlear implant as in Australia, such children would have received a cochlear implant.

7.3 Future Studies

Future developments with the test include its administration to a large group of hearing impaired children and standardisation using a recorded version.

The benefit of using the recorded format is to ensure that the test material presented is always the same and to avoid exaggerated speech patterns that can be confusing. For example, a speaker could exaggerate their pronunciation, so that the words that normally have one beat for hearing impaired children, such as *rumah* became something like *ru-mah*, and this form of the words has two beats to the hearing impaired children. Thus, this will interfere with the consistency of the test results. These children are more likely to have people exaggerate the speech pattern, because they have a problem with their hearing. However, stress pattern would not be a problem in live voice presentation because stress is not relevant in Indonesian. Another benefit of using a recorded format is that we would be able to obtain separate ear information.

It is intended that in the future, as the development of the tests in this study progresses, a more detailed suprasegmental assessment and segmental assessment will be developed. INDO-SPASP test has a hierarchical order of difficulty, which is similar to the situation in tests such as GASP! (Erber, 1982, 2011) and PLOTT (Plant & Westcott, 1983), which are described below.

GASP! was designed to describe a child's auditory abilities using three different stimulus types (phonemes, words, and sentences) and three different response types (detection, identification, and comprehension). The standard procedure involves closed-set administration of the word identification subtest and open set presentation of the sentence portion of GASP!.

The PLOTT consists of nine subtests of increasing difficulty ranging from a relatively simple detection level task to more difficult discriminations involving the perception of spectral information. The aim of the test is to provide a comprehensive view of a hearing-impaired child's ability to perceive both suprasegmental and segmental speech features.

It is possible that THRIFT (Boothroyd, 1991) which was mentioned in chapter Introduction to INDO-SPASP, may also be considered for use, (with modifications) especially in the syllable pattern and the balance between vowel contrasts and consonant contrasts, as part of the INDO-SPASP test battery at a later stage (as a more advanced/detailed segmental test).

It is intended that INDO-SPASP in the future will provide the Indonesian hearing care professional with a detailed overview of the individual children's speech perception capabilities and will:

- distinguish between those children who can perceive the spectral/segmental components and those who can perceive only suprasegmental components;

- have a hierarchical order of difficulty. Hence, the test will consist of several subtests which will start from detection as the most basic auditory skill to the most complex auditory skill, that is, comprehension.

At this stage, the test developed in this study has two subtests only: (i) detection; and (ii) identification that need minimal investment in technology, be easy to learn for those who administering it; and have a variety of practical applications.

CHAPTER 8: CONCLUSIONS AND OVERALL DISCUSSION

In Chapter 8, INDO-SPRITT and INDO-SPASP will be discussed separately. The design of the speech perception test in BI is basically similar to other tests in English, such as NU-CHIPS (Elliot & Katz, 1980) or MTS (Erber & Witt, 1977). However, there is a difference in the construction of the material that is influenced by language constraints.

8.1 INDO-SPRITT

INDO-SPRITT is the first standardised speech audiometry for children in Indonesia. The present data indicate the following:

- The INDO-SPRITT material is familiar to normally hearing children as young as 4.5 years old and is familiar to severely to profound hearing impaired children as young as 7.6 years old.
- The reliability and validity of the INDO-SPRITT test is adequate for children as young as 4.5 years old.
- The administration of 5 reversals is enough to estimate SRT without sacrificing the reliability of the threshold measurement. It is important to note that the start of counting the 5 reversals should be made after first allowing three practice reversals.
- The normative reference value of SRT ranges from 6.87 to 25.99 dBHL.
- The age effect in the SRT of 4-5 year old group was significantly higher than SRT of the 8-9 year old and 10-13 year old groups. The SRT of the 6-7 year old group and SRT of 8-9 year old group were significantly higher than SRT of the 10-13 year old group.

- The correlation between SRT and 3 FAHL is modest but this should be useful in a clinical setting, because it indicates that increasing hearing loss is associated with poorer speech recognition thresholds.

- The formula that will be used in the clinical practice to predict SRT and 3 FAHL using 5 reversals for hearing ≤ 20 dBHL is as follows::

$$3 \text{ FAHL (in dBHL)} = 8 + (0.33 \times \text{SRT}) \pm 10$$
$$\text{SRT (in dBHL)} = 10.5 + (0.5 \times 3 \text{ FAHL}) \pm 10$$

- INDO-SPRITT is to be used for children who use BI for their daily communication or as their first language, as well as for children with normal to moderately-severe hearing loss.

The characteristics of INDO-SPRITT material are as follows:

- It uses bisyllabic words of BI and representable by pictures.

- It has lists and items with equal difficulty and phonemically balanced.

- It uses a picture pointing task in a closed response set.

- It has foil-minimal pair or contrast in an analogous environment.

- It has recorded material with a native speaker who uses BI for daily communication.

- It is to be administered in a quiet environment, using an adaptive technique and a whole word scoring method.

It is commonly known that word familiarity or vocabulary restriction is an important consideration in constructing a speech perception test for children due to its age effect. However, in this study it was found that age has an effect on the perception or identification of pictures. Little is known or discussed in the audiology literature about the importance of how pictures are perceived in the construction of speech tests for young children. Thus, it is important to consider both the developmental word recognition and picture effects in the constructing speech tests that use pictures for this population.

It should also be stressed that the foils of test items are an important factor in developing a closed set speech perception test, because they may affect the equal difficulty between test items and lists. A lack of equal difficulty between test items and lists may influence the reliability of the test and hence influence the sensitivity of the test being constructed (Bamford & Wilson, 1979; Dillon, 1983). Variation of foils in the INDO-SPRITT are: (i) two test items that each has only one effective foil; (ii) six test items that each has two effective foils; and (iii) 42 test items that each has three effective foils. To compare NU-CHIPS with INDO-SPRITT materials, INDO-SPRITT have more phonologically consistent foils (Section 3.2.3). Furthermore, INDO-SPRITT has a clearer equal difficulty between lists than NU-CHIPS, because there is the same number of foils for a given test item in different lists. Somewhat surprisingly is in NU-CHIPS, the number of effective foils for some test items differs between lists.

The present study also indicates that INDO-SPRITT is a sensitive test, as the following have been established (Bilger, 1984; Elkins, 1984):

- its purpose
- its target population
- the validity and reliability of the test
- a procedure for test administration, scoring and interpretation.

Furthermore INDO-SPRITT has been designed to limit the influence of receptive language ability on test performance, as well as the effect of extra-auditory (cognitive) factors such as intelligence and attentional memory on children’s performance

A potential weakness of the test in the study is the absence of normative data from the hearing impaired subjects. This is because, at the time of this study, the majority of the hearing impaired children in the hearing institution in Indonesia had a severely to profound sensorineural loss. However, this issue has been overcome, to some extent, by establishing the normative value from the normal hearing children. The mean SRTs in this study provide a normative reference against which SRTs of hearing-impaired individuals can be compared. Thus, INDO-SPRITT can be used:

- To complement pure tone audiometry
- To provide unaided speech threshold
- To cross-check pure tone audiometry results
- To demonstrate the impact of hearing loss on speech perception to parents and teachers
- To provide aided speech thresholds for patients with hearing aids or cochlear implants
- To provide a method of demonstrating aided advantages to parents and teachers

It is also realised that the lack of a formal hearing test to identify subjects with normal hearing is a weakness, as some of the subjects may actually have had a hearing loss which was possibly not great enough to attract the attention of those involved with them. Another problem is the lack of acoustically treated rooms in which to carry out the study. The ambient noise levels may have affected the results, but, a sound level meter was used to measure sound levels and the test was halted if ambient noise if the ambient noise levels exceeded 55 dBA (as mentioned on page 55 last paragraphs).

The small numbers of subject involved in this study might be described as a weakness, but, the numbers were sufficiently large to provide useful data. Further studies might be useful to confirm the findings of the present study.

The study on INDO-SPRITT led to the development of INDO-SPASP because it was found that the majority of children in the hearing institution in Indonesia had severe to profound hearing loss. There was no data available at the time of study regarding the population of hearing impaired children with different degrees of hearing loss. This lack of statistical data in Indonesia, which is also found in many other developing countries, is consistent with previous studies about audiology in developing countries (WHO, 2004; McPherson & Olusanya 2008; Smith, 2008).

8.2 INDO-SPASP

It is intended that INDO-SPASP will be used as a standardised speech audiometry for children in Indonesia. The results obtained in this pilot study indicate that the speech test provides much useful information on the speech perception capabilities of severely and profoundly hearing-impaired children.

At this stage, INDO-SPASP consists of two subtests: (i) phoneme detection; and (ii) word identification.

8.2.1 Subtest 1 – phoneme detection

The phoneme detection subtest follows the Ling Sound Test (Ling, 1976) as the basis of a detection test. The Ling Sound Test was originally developed for the North American population. The differences in production and spectral content of North American and Indonesian vowels were investigated to determine if Ling sounds were appropriate for the Indonesian population or whether the original Ling Sound Test needed to be modified, as in the case with the Australian population. The results of the investigation showed that the spectral contents of Indonesian vowels are similar to the spectral content of the North American vowels compared to the spectral content of Australian vowels. Therefore, the phonemes used in the INDO-SPASP are the same as in the original Ling six sound test which are /a/, /i/, /u/, /m/, /ʃ/ and /s/.

Subtest 1 (phoneme detection) was important in this study because it helped to determine whether it was reasonable to present the Subtest 2 (word identification). Furthermore, it established an indication of the audibility of various frequencies that children could hear. In future, a clinical practice subtest can be used for the following purposes:

- To determine if a child's hearing aid is adequate for auditory speech perception, providing enough gain and output level so that most phonemes can be detected.

- To establish whether the teacher's voice is heard by the child at a sufficiently high enough level for auditory communication.

- To establish whether it is reasonable to present the second subtest. This avoids the situation where the person administering the test is attempting to test a child's speech perception abilities with inaudible sound.

It should be stressed that the philosophy behind the detection test using phonemes is to determine whether certain frequencies are perceived and does not focus specifically on identifying vowels or consonants (Ling, 1976).

8.2.2 Subtest 2 – word identification.

Preliminary results of the pilot study on Subtest 2 (word identification) indicate the following:

- The identification subtest of INDO-SPASP appeared to be able to distinguish between children who can perceive spectral/segmental components and those who can perceive only suprasegmental components.

- Children with a hearing loss over 91 dBHL identify words primarily on the basis of their syllable/suprasegmental pattern, whereas children with a less profound hearing loss can use segmental cues.

- Children with hearing loss greater than 91 dB demonstrate a range of speech perception abilities. This is consistent with previous studies (Cramer & Erber, 1974; Erber, 1974; Risberg et al., 1975; Erber, 1979, 1980; Plant, 1984) that audiogram is not a good predictor of speech perception ability for profound hearing loss.

The results from the word identification assessment can be used:

- To assist in educational placement
- To establish the objectives of auditory training
- To measure the effects of auditory training
- To assist in selecting a listening device for the child.

It is hoped that the collection of data from these diagnostic tools should, not only help hearing impaired children, but also help to increase the awareness of parents, teachers and the general public on deafness and its impact, which will lead to a stronger advocacy.

Further study on developing speech audiometry in other ethnic languages is highly desirable since Indonesia has more than 700 living languages spoken across the Indonesian archipelago (Lewis, 2009). It is important to note that an estimated 23 million people speak BI as a first language and an additional 140 million speak it as a second language (Grimes, 1996; Gordon, 2005).

REFERENCES

Agung, K. B., Purdy, S. C. & Kitamura, C. (2005). The Ling Sound Test Revisited. *Australian and New Zealand Journal of Audiology, 27*(1), 33-41.

American Speech-Language Hearing Association. (1988). Guidelines for determining the threshold level for speech. *Asha, 30*(3), 85-89.

Asher, J. W. (1958). Intelligibility tests: A review of their standardization. Some experiments and a new test. *Speech Monographs, 25*(1), 14-28.

Ashoor, A. A. & Prochzka, T. (1985). Saudi Arabic speech audiometry for children. *British Journal of Audiology, 19*(3), 229-238.

Australian Hearing Manual of Speech Perception. (1999). *The Ling five and seven sound test*. Sydney: Australian Hearing.

Bamford, J. & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (eds), *Speech-hearing tests and the spoken language of hearing-impaired children*. London: Academic Press.

Bench, J. & Bamford, J. (1979). Speech-hearing tests and the spoken language of hearing-impaired children. London: Academic Press.

Bernstein, R. S. & Grave, J. S. (1990). A Method for Determining Hearing Sensitivity in Infants: The Interweaving Staircase Procedure (ISP). *Journal of the American Academy of Audiology, 1*(3), 138-145.

Bilger, R. C. (1984). Speech recognition test development. In E. Elkins (ed.), *Speech recognition by the hearing impaired. ASHA Reports*, 14, 2-15. Rockville: ASHA.

Bilger, R. C. & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *Journal of Speech, Language and Hearing Research*, 19(4), 718-748.

Black, J. W. (1957). Multiple-choice intelligibility tests. *Journal of Speech and Hearing Disorders, 22*, 213-235.

Black, J. W. (1968). Responses to multiple-choice intelligibility tests. *Journal of Speech Hearing Research, 11*, 453-466.

Blamey, P. J. (2003). Development of Spoken Language by Deaf Children. In M. Marschark & P. E. Spencer (eds), *The Oxford Handbook* of *Deaf Studies, Language, And Education*. New York: Oxford University Press.

Bode, D. L. & Carhart, R. (1973). Measurement of articulation functions using adaptive test procedures. *IEEE Transactions on Audio and Electroacoustics, AU-21*(3), 196-201.

Bode, D. L. & Carhart, R. (1974). Stability and accuracy of adaptive tests of speech discrimination. *Journal Acoustical Society of America, 56*(3), 963-970.

Boe, L. J., Heim, J. L., Honda, K. & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, 30(3), 465-484

Boothroyd, A. (1968). Developments in Speech Audiometry. *Sound, 2*(1), 3-10.

Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 27, 134-144.

Boothroyd, A. (1988). Perception of speech pattern contrasts from auditory presentation of voice fundamental frequency. *Ear And Hearing*, 9(6), 313-321.

Boothroyd, A. (1991). Assessment of speech perception capacity in profoundly deaf children. *American Journal of Otology*, 12 (supplement), 67-72.

Bradley, J., Lancashire, I., Presutti, L. & Stairs, M. (1995). *Text Analysis Computing Tools (TACT)*, version 2.1.4, from http://projects.chass.utoronto.ca/tact/

Brandy, W. T. (1966). Reliability of voice tests of speech discrimination. *Journal of Speech, Language and Hearing Research, 9*(3), 461-465.

Brandy, W. T. (2002). Speech Audiometry. In J. Katz (ed.), *Handbook of Clinical Audiology*. Philadelphia: Lippincott Williams & Wilkins.

Branson, J. & Miller, D. (2002). Damned for Their Difference: The Cultural Construction of Deaf People as Disabled: A Sociological History. Washington DC: Gallaudet University Press.

Breakey, M. R. & Davis, H. (1949). Comparisons of thresholds for speech: Word and Sentence tests; Receiver vs Field, and Monoaural vs Binaural Listening. *Laryngoscope, 59*(3), 236-250.

Burns, R. B. (1994). *Research methods*. Australia: Longman.

Byrne, D. (1983). Word familiarity in speech perception testing of children. *Australian Journal of Audiology, 5*(2), 77-80.

Cacace, A. T. (1998). Central auditory processing disorder in school-aged children: A critical review. *Journal of Speech, Language and Hearing Research, 41*(2), 355-373.

Campanelli, P. (1962). A measure of intra-list stability of four PAL word lists. *Journal of Audiology Research, 2*, 50-55.

Campbell, R. A. (1965). Discrimination test word difficulty. *Journal of the Speech and Hearing Research, 8*(1), 13-22.

Carhart, R. (1971). Observations on relations between thresholds for pure tones and for speech. *Journal of Speech and Hearing Disorder, 36*(4), 476-483.

Carhart, R. & Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone Thresholds. *Journal of Speech and Hearing Disorder, 24*(4), 330-345.

Carney, A. E. & Moeller, M. P. (1998). Treatment efficacy: Hearing loss in children. *Journal of Speech, Language & Hearing Research, 41*(1), S61-S84.

Ching, T.Y.C., Leigh, G. & Dillon, H. (2013). Introduction to the Longitudinal Outcomes of Children with Hearing Impairment (LOCHI) study: Background, design, sample characteristics. *International Journal of Audiology, 52*(Supplement), S4-S9.

Cole, R. A. & Jakimik, J. (1980). How are syllables used to recognize words? *Journal Acoustical Society of America, 67*(3), 965-970.

Cowie, R. & Douglas-Cowie, E. (1992). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Vol. 62 of Trend in Linguistics. Studies and Monongraphs (TiLSM)). Berlin: Walter de Gruyter.

Cox, F. (1999). Vowel change in Australian English. *Phonetica, 56*, 1-27.

Cox, F. (2008). The acoustic characteristics of nasals. Retrieved 1 September 2014, from clas.mq.edu.au/speech/acoustics/consonants/nasalweb.html

Cox, F. (2012). *Australian English Pronunciation and Transcription*. Cambridge: Cambridge University Press

Cramer, K. D. & Erber, N. P. (1974). A spondee recognition test for young hearing-impaired children. *Journal of Speech and Hearing Disorders, 39*(3), 304-311.

Crandell, C. C. (1991). Individual differences in speech recognition ability: Implications for hearing aid selection. *Ear & Hearing, 12*(6), 100S-108S.

Dardjowidjojo, S. (1978). *Sentence Patterns of Indonesian*. Honolulu: Honolulu University of Hawaii Press.

Davis, H. & Hudgins, C. (1946). The selection of hearing aids. *Laryngoscope, 56*, 85-146.

Dillon, H. & Ching, T. (1995). What Makes a Good Speech Test? In G. Plant & K. E. Spens (eds), *Profound Deafness and Speech Communication*. London: Whurr Publishers.

Dillon, H. (1983). The effect of test difficulty on the sensitivity of speech discrimination tests. *Journal of the Acoustical Society of America, 73*(1), 336-344.

Dirks, D. D., Morgan, D. E. & Dubno, J. R. (1982). Procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorder, 47*(2), 114-123.

Dixon, W. J. & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistics Association, 43,* 109-126.

Dobie, R. A. & Hemel, S. V. (2004). *Hearing Loss: Determining Eligibility for Social Security Benefits*. Washington DC: National Academies Press (US).

Doyne, M. P. & Steer, M. D. (1951). Studies in speech reception testing. *Journal of Speech and Hearing Disorder, 16*(2), 132-139.

Access Economics. (2006). Listen, Hear! The economic impact and cost of hearing loss in Australia. Melbourne: Access Economics.

Egan, J. P. (1948), Articulation testing methods. *Laryngoscope*, 58, 955-981.

Elkins, E. (1984). Recommendations for speech recognition research. In E. Elkins (ed.), *Speech Recognition by the hearing impaired*, 14, 77-79). Rockville: American Speech-Language-hearing Association.

Elliot, L. L. & Katz, D. R. (1980). Development of a new children's test of speech discrimination. St Louis: Auditec.

Elpern, B. S. (1961). The relative stability of half-list and full-list discrimination tests. *The Laryngoscope, 71*(1), 30-36.

Erber, N. P. (1972). Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research, 15*(2), 413-422.

Erber, N. P. (1974). Pure-tone thresholds and word recognition abilities of hearing-impaired children. *Journal of Speech and Hearing Research, 17*(2), 194-202.

Erber, N. P. (1979). Speech perception by profoundly hearing impaired children. *Journal of Speech and Hearing Disorder, 44*(3), 255-270.

Erber, N. P. (1980). Use of the auditory numbers test to evaluate speech perception abilities of hearing impaired Children. *Journal of Speech and Hearing Disorder, 45*(4), 527-532.

Erber, N. P. (1982). *Auditory training*. Washington: Alexander Graham Bell Association for the Deaf.

Erber, N. P. (2011). *Auditory communication for deaf children*. Camberwell: ACER Press.

Erber, N. P. & Alencewicz, C. M. (1976). Audiologic evaluation of deaf children. *Journal of Speech and Hearing Disorder, 41*(2), 256-267.

Erber, N. P. & Witt, L. H. (1977). Effects of stimulus intensity on speech perception by deaf children. *Journal of Speech and Hearing Disorders, 42*(2), 271-278.

Fairbanks, G. & Miron, M. S. (1957). Effects of vocal effort upon the consonant-vowel ration within the syllable. *Journal of Acoustical Society of America, 29*(5), 621-626.

Finitzo-Hieber, T., Gerling, I. J., Matkin, N. D. & Cherow-Skalka, E. (1980). A sound effects recognition test for the pediatric audiological evaluation. *Ear & Hearing, 1*(5), 271-276.

Fletcher, H. (1929). *Speech and Hearing*. New York: Van Nostrand.

Fletcher, H. (1950). A method of calculating hearing loss for speech from an audiogram. *Journal of the Acoustical Society of America, 22*(1), 1-5.

Geers, A. E. & Moog, J. S. (1989). Factors predictive of the development of literacy in profoundly hearing impaired adolescent. *Volta Review, 91*, 69-86.

Gordon, R. G. (2005). *Ethnologue: Languages of the world*. Dallas: SIL International.

Grimes, C. E. (1996). Indonesian-the official language of a multilingual nation. In S. A. Wurm, P. Muhlhausler & D. T. Tyron (eds), *Atlas of Languages of Intercultural communication in the Pacific, Asia, and the Americas (Trends in Linguistics. Documentation 13)*, pp. 719-727. Berlin: Mouton de Gruyter.

Grubb, P. (1963). Some considerations in the use of half-list speech discrimination tests *Journal of Speech and Hearing Research, 6*(3), 291-297.

Hack, Z. C. & Erber, N. P. (1982). Auditory, visual, and auditory – visual perception of vowels by hearing-impaired children. *Journal of Speech and Hearing Research, 25*(1), 100-107.

Hagerman, D. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology, 11*, 79-87.

Hagerman, D. (1984). Clinical measurements of speech reception threshold in noise. *Scandinavian Audiology, 13*, 57-63.

Hardy, W. G. & Bordley, J. E. (1951). Special techniques of testing the hearing of children. *Journal of Speech and Hearing Disorders, 16*(2), 122-131.

Harrington, J., & Cassidy, S. (1999). Techniques in Speech Acoustics. United States: Kluwer Academic Publishers.

Henderson, I. (2006). *East Java Early Childhood Needs Assessment*. Patricia O'Sullivan Humanitarian Project.

Henkel, E. (2005). *Handbook of research in second language teaching and learning*. New Jersey: Lawrence Erlbaum Associates.

Heuven, V.J. & Zanten, E. (1997). *Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian.* Paper presented at the 7th International Conference on Austronesian Linguistics, Amsterdam/Atlanta.

Hirsh, I. J., Reynolds, E. G. & Joseph, M. (1954). Intelligibility of different speech materials. *Journal of the Acoustical Society of America, 26*(4), 530-538.

Hughson, W. & Thompson, E. (1942). Correlation of hearing acuity for speech with discrete frequency audiograms method of determination. *Archives of Otolaryngology, 36*(4), 526-540.

Hughson, W. & Westlake, H. D. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions of the American Academy of Ophthalmology and Otolaryngology, 48*(Suppl.), 1-15.

International Organization for Standardization 8253-3, (2012). Acoustic-Audiometric test methods. Part 3 Speech audiometry.

Jerger, S. (1984). Speech Audiometry. In J. Jerger (ed.), *Pediatric Audiology*. San Diego: College Hill Press.

Jerger, S., Lewis, S., Hawkins, J. & Jerger, J. (1980). Pediatric Speech Intellibility Test I. Generation of Test Material. *International Journal of Pediatric Otorhinolaryngology, 2*(3), 217-230.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*(3), 227-229.

Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics, 63*(8), 1377-1388.

Kei, J., Chan, T., Ng, M.C.M.L., Lowe, C., Lai, D. M., Ng, Y. H. & Kwan, E. (1991). Cantonese speech audiometry for children of Hong Kong. *Australian Journal of Audiology, 13*(2), 41-45.

Kendall, D. C. (1954). Audiometry for young children. *The Teacher of The Deaf, LII/52*(307), 18-23.

Kennedy, C. R., Campbell, M. J., Law, C. M., Mullee, M., Petrou, S., Watkin, P. & Stevenson, J. (2006). Language ability after early detection of permanent childhood hearing impairment. *New England Journal of Medicine*, *354*, 2131-2141.

Kirk, K. I., Diendorf, A. O., Pisoni, D. B. & Robbins, A. M. (1997). Assessing speech perception in children. In L. L. Mendel & J. L. Danhauer (eds), *Audiologic Evaluation and Management and Speech Perception Assessment*, pp. 101-131. San Diego; London: Singular Publishing Group.

Kosky, C. & Boothroyd, A. (2003). Validation of an on-line implementation of the imitative test of speech pattern contrast Perception (IMSPAC). *Journal of the American Academy of Audiology, 14*(2), 72-83.

Kruger, B. & Mazor, R. M. (1987). Speech audiometry in the USA. In M. Martin (ed.), *Speech Audiometry*, pp. 207-235. London; New York; Philadelphia: Taylor & Francis.

Kumar, S. (2001). WHO tackles hearing disabilities in developing world. *The Lancet, 358*, 219.

Ladefoged, P. & Johnson, K. (2011). *A course in phonetics*. Boston: Wadsworth, Cengage.

Ladefoged, P. (1993). *A course in phonetics*. Sydney: Harcourt Brace Jovanovich.

Lane, H. L. (2006). *The Deaf Experience: Classics in Language and Education* (F. Philip, trans. vol. 5). Chicago, Illinois: Gallaudet University Press.

Lane, H. L. (1984). *The Deaf Experience: Classics in Language and Education* (F. Philip, Trans.). Cambridge, Massachusetts: Harvard University Press.

Lapoliwa, H. (1981). *A generative approach to the phonology of Bahasa Indonesia*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

Laurence, R. F., Moore, B.C.J. & Glasberg, B. R. (1983). A comparison of behind-the-ear high-fidelity linear hearing aids and two channel compression ands, in the laboratory in everyday life. Br. I. Audioi. *17*, 3148.

Lawson, G. & Peterson, M. (2011). *Speech Audiometry*. San Diego: Plural Publishing.

Lehiste, I. & Peterson, G. E. (1959). Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustical Society of America, 31*(3), 280-286.

Leigh, G., Newall, J. P. & Newall, A. T. (2010). Newborn Screening and Earlier Intevention with Deaf Children: Issues for the Developing World. In M. Marschark & P. E. Spencer (eds), *The Oxford Handbook of Deaf Studies, Language, and Education* (vol. 2). New York: Oxford University Press.

Levitt, H. & Rabiner, L. R. (1967). Use of a sequential strategy in intelligibility testing. *Journal Acoustical Society of America, 42*(3), 609-612.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal Acoustical Society of America, 49*(2B), 467-477.

Levitt, H. (1978b). Adaptive testing in audiology. *Scandinavian Audiology, Supplement 6*, 241-291.

Lewis, M. P. (2009). Ethnologue: Languages of the World. Dallas: SIL International.

Ling, D. (1976). *Speech and the hearing-impaired child: Theory and practice*. Washington: Alexander Graham Bell Association for the Deaf.

Ling, D. (2002). *The Ling-six sound test.* Paper presented at the 2002 Alexander Graham Bell Convention St. Louis, MO.

Lovegrove, R., Dillon, H. & Mackie, K. (1991). North Western University Children's Perception of Speech (NU-CHIPS) Audiology Circular, vol. 4, pp. 1-8. Australia: NAL.

Lyregaard, P. (1987). Towards a theory of speech audiometry tests. In M. Martin (ed.), *Speech Audiometry*, p. 33. London; New York: Taylor & Francis.

Mackersie, C. L. (2002). Tests of speech perception abilities. *Current Opinion in Otolaryngology & Head and Neck Surgery, 10*, 392-397.

Mackie, K. & Dermody, P. (1986). Use of a monosyllabic adaptive speech test (MAST) with young children. *Journal of Speech and Hearing Research, 29*(2), 275-281.

Mackie, K. C. & Dermody, P. J. (1982). Word intelligibility tests in audiology for the assessment of communication adequacy. Canberra: National Acoustic Laboratory.

Maltby, M. (2000). A new speech perception test for profoundly deaf children. *Deafness and Education International, 2*(2), 86-100.

Margolis, R. H. & Millin, J. P. (1971). An item-difficulty based speech discrimination test. *Journal of Speech and Hearing Research, 14*(4), 865-873.

Markides, A. (1980). The relation between hearing loss for pure tones and hearing loss for speech among hearing-impaired children. *British Journal of Audiology, 14*(4), 115-121.

Markides, A. (1987). Speech tests of hearing for children. In M. Martin (ed.), *Speech Audiometry*, pp. 155-170. London; New York; Philadelphia: Taylor & Francis.

Martin, F. N. & Clark, J. G. (2003). *Introduction to Audiology*. Boston: Allyn & Bacon.

Martin, F. N., Champlin, C. A. & D.Perez, D. (2000). The question of phonetic balance in word recognition testing. *Journal of the American Academy of Audiology, 11*(9), 489-493.

Martony, J. (1974). Some psychoacoustic tests with hearing impaired children. *STL-QPSR, 15*(2-3), 72-89.

McArdle, R. & Hnath-Chisolm, T. (2009). Speech audiometry. In J. Katz, L. Medwetsky, R. F. Burkard & L. Hood (eds), *Handbook of Clinical Audiology*, pp. 64-79. Philadelphia: Lippincot Williams & Wilkins.

McFarlane, D. (1940). Speech hearing and speech interpretation testing. *Oto-Laryngology, 31*, 517-528.

McPherson, B. & Olusanya, B. O. (2008). Screening for hearing loss in developing countries. In B. McPherson & R. Brouillette (eds), *Audiology in Developing Countries*, pp. 75-105. New York: Nova Science Publishers.

McPherson, B. (2008). Audiology: A Developing Country Context. In B. McPherson & R. Brouillette (eds), *Audiology in Developing Countries*, pp. 5-20. New York: Nova Science Publishers.

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology, 47*(9), 546-553.

Mendel, L. L. & Danhauer, J. L. (1997). Test Development and Standardization. In L. L. Mendel & J. L. Danhauer (eds), *Speech Perception Assessment*, pp. 7-15. San Diego; London: Singular Publishing Group.

Merklein, R. A. (1981). A short speech perception test. *The Volta Review, 83*(1), 36-45.

Miller, G.A., Heise, G.A. & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology 41*, 329-335.

Moeller, M. P., Carr, G., Seaver, L., Stredler-Brown, A. & Holzinger, D. (2013). Best Practices in family-centered early intervention for children who are deaf or hard of hearing: An international consensus statement. *Journal of Deaf Studies and Deaf Education, 18*, 429-445.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics, 106*(3), e43.

Moog, J. S. & Geers, A. E. (1990). *Early Speech Perception Test for Profoundly Hearing-Impaired Children*. St. Louis: Central Institute for the Deaf.

Moore, R. C. (2014, 1 July). [Personal Communication].

Mukari, S. Z. M. & Abdul Hamid, B. (2008). Penilaian Respons Auditori kepada Pertuturan. Bangi: Universiti Kebangsaan Malaysia

Nilsson, M., Soli, S. D. & Sullivan, J. A. (1994). Development of the hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal Acoustical Society of America, 95*(2), 1085-1099.

Northern, J. L. & Downs, M. P. (2002). Hearing and Hearing Loss in Children. In J. L. Northern & M. P. Downs (Eds), *Hearing in Children*, pp. 1-28. Baltimore: Lippincott Williams & Wilkins.

Nunnally. (1978). *Psychometric Theory*. New York: McGraw-Hill.

Odé, C. (1994) On the perception of prominence in Indonesian. in: C. Odé and V. J. van Heuven (eds), *Experimental studies of Indonesian prosody,* pp. 27-107, Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

Olsen, W. O. & Matkin, N. D. (1991). Speech Audiometry. In W. F. Rintelmann (ed.), *Hearing Assessment*. Austin: Pro-Ed.

Olusanya, B. O. & Newton, V. E. (2007). Global burden of childhood hearing impairment and disease control priorities for developing countries. *The Lancet, 369*, 1314-1317.

Olusanya, B. O., Luxon, L. M. & Wirz, S. L. (2005a). Childhood deafness poses problems in developing countries. British Medical Journal, 330, 480-481.

Olusanya, B. O. (2005b). Can the world's infants with hearing loss wait? *International Journal of Pediatric Othorinolaryngology, 69*, 735-738.

Owens, E. (1961). Intelligibility of words varying in familiarity. *Journal of Speech and Hearing Research, 4*(2), 113-129.

Palmer, J. M. (1955). The effect of Speaker differences on the intelligibility of PB word lists. *Journal of Speech Hearing Disorders, 20*(2), 192 - 195.

Pickett, J. M., Martin, E. S., Johnson, D., Brandsmith, S., Daniel, Z., Willis, D. & Otis, W. (1972). On patterns of speech feature reception by deaf listeners. . In G. Fant (ed.), *Speech communication ability and profound deafness*. Washington: A. G. Bell Association for the Deaf.

Plant, G. (1984). A diagnostic speech test for severely and profoundly hearing-impaired children. *Australian Journal of Audiology*, *6*(1), 1-9.

Plant, G. (1990). The development of speech tests in Aboriginal languages. *Australian Journal of Audiology, 13*(1), 30-40.

Plant, G. & Westcott, S. (1983). *The PLOTT Test*. Chatswood: National Acoustic Laboratories.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America, 63*(2), 533-549.

Plomp, R. (1986). A signal-to noise ration model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research, 29*(2), 146-154.

Plomp, R. & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18*(1), 43-52.

Raffin, M. & Thornton, A. (1977). *A model for estimating the variability across equivalent test forms*. Paper presented at the Annual Convention of Canadian Speech and Hearing Association, Victoria.

Resnick, D. M. (1962). Reliability of the 25 word phonetically balanced lists. *Journal of Audiology Research, 2*, 5-12.

Risberg, A. (1976). Diagnostic rhyme test for speech audiometry with severely hard of hearing and profoundly deaf children. *STL-QPSR, 17*(2-3), 40-58.

Roosman, L. M. (2004). Non-native accents in Dutch word stress realisation. *Leiden Working Papers in Linguistics 1.1,* 63-81.

Ross, M. & Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13*(1), 44-53.

Ross, M. & Lerman, J. (1970). *Word Intelligibility by Picture Identification.* Canada: J. M. Dent & Sons.

Ruscetta, M. N., Palmer, C. V., Durrant, J. D., Grayhack, J. & Ryan, C. (2005). Validity, internal consistency and test/retest reliability of a localization disabilities and handicaps questionnaire. *Journal of the American Academy of Audiology, 16*(8), 585-595.

Sacks, O. (1991). *Seeing voices*. London: Pan Books.

Shutts, R. E., Burke, K. S. & Creston, J. E. (1964). Derivation of twenty-five word PB lists. *Journal of Speech and Hearing Disorder, 29*, 442-447.

Siegenthaler, B. M. & G., H. (1966). Development of two standardized measures of hearing for speech by children: Cooperative Research Program. Washington: Department of Health, Education and Welfare.

Silman, S. & Silverman, C. A. (1997). *Auditory diagnosis: Principals and applications.* San Diego: Singular Publishing Group.

Silverstein, B., Bilger, R. C., Hanley, T. D., & Steer, M. D. (1953). The relative intelligibility of male and female talkers. *Journal of Educational Psychology, 44*(7), 418-428.

Small, A. M. (1973). Psychoacoustics. In F. D. Minifie, T. J. Hixon & F. Williams (eds), *Normal Aspects of Speech, Hearing, and Language*, pp. 347). Englewood Cliffs: Prentice-Hall.

Smith, A. W. (2008). Demographics of hearing loss in developing countries. In B. McPherson & R. Brouillette (eds), *Audiology in Developing Countries*, pp. 21-50). New York: Nova Science Publishers.

Soderberg, C. D. & Olson, K. S. (2008). Indonesian. Journal of the International Phonetic Association: Illustrations of the IPA, 38(2), 209-213.

Soewito, A., Djoko S. S. & Soejarno, A. H. (1984). Pembakuan Tes Bisik Kata Bisilabik Bahasa Indonesia Untuk 'Screening' Pendengaran Anak Sekolah. Yogyakarta: Universitas Gajah Mada.

Soli, S. D. (2008). Some thoughts on communication handicap and hearing impairment. *International Journal of Audiology, 47*(6), 285-286.

Stach, B. A. (1998). Speech Audiometry. In B. A. Stach (ed.), *Clinical Audiology, An Introduction*, pp. 229-252. San Diego; London: Singular Publishing Group.

Steinberg, W. J. (2011). *Statistics Alive* (2nd ed.). United States of America: SAGE Publications, Inc.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M. & Lewis, D. E. (2001). Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. *Journal of the Acoustical Society of America, 110*(4), 2183-2190.

"Success for kids with hearing loss". Retrieved 20 July, 2014, from http://successforkidswithhearingloss.com/relationship-hl-listen-learn

Tye-Murray, N. (2009). *Foundations of aural rehabilitation*. St. Louis: Delmar, Cengage Learning.

Walden, B. E. & Montgomery, A. A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. *Journal of Speech and Hearing Research, 18*(3), 444-455.

Walden, B. E. (1984). Validity issues in speech recognition test development. In E. Elkins (ed.), *Speech recognition by the hearing impaired. ASHA reports*, *14*, 16-18. Rockville: ASHA.

Wang, B. K. (2006). Let's Get Real! Understanding the distinctions between aided SRT and WRS. *The Hearing Review*.

Watson, T. J. (1957). Speech audiometry in children. In A.W.G. Ewing (ed.), *Educational Guidance and the Deaf Child*. Manchester: Manchester University Press.

Wechsler, D. (2002). The Wechsler Preschool and Primary Scale of Intelligence, (WPPSI-III). San Antonio: The Psychological Corporation.

World Health Organisation (2005). Prevention of deafness and hearing impairment. World Health Organisation.

World Health Organisation Regional Office for South East Asia region (2004). State of hearing and ear care in the South East Asia region. New Delhi WHO Regional Office for South East Asia.

Wright, D. (1990). *Deafness*. London: Faber & Faber.

World Population Review (2014). Indonesia Population 2014. Retrieved 3 December 2014, from WPR website: http://worldpopulationreview.com/countries/indonesia-population/

Yoshinaga-Itano, C. (2003a). Sensitive Periods for Access to Language: The Remarkable Infant with Hearing Loss. Paper presented at the Danavox Symposium.

Yoshinaga-Itano, C. (2003b). Early Intervention After Universal Neonatal Hearing screening: Impact on Outcomes. *Mental Retardation and Developmental Disabilities Research Reviews 9*(4), 252-266.

Zanten, E. V. (1989). The Indonesian vowels: Acoustic and perceptual exploration. Leiden: University of Leiden.

Zanten, E. V. & Goedemans, R. (2009). Prominence in Indonesian; Stress, phrases and boundaries. *Wacana, 11*(2), 197-225.

Zanten, E. V. & Heuven, V.J.V. (1998). Word stress in Indonesian: Its communicative relevance. *Bijdragen tot de Taal-, Land-en Volkenkunde, 154*(1), 129-149.

Zanten, E. V. & Heuven, V.J.V. (2004). Word stress in Indonesian: Fixed or free? In B. K. Purwo (ed.), *Studies in Malay Dialects Part IV (Linguistic Studies of Indonesian and Other Languages in Indonesia)*, *53*, 1-20. Jakarta: NUSA.

List of Appendices

APPENDIX A:
VARIATION OF FOILS IN NU-CHIPS (AMERICAN ENGLISH)

Pronunciation or phoneme was checked using Collins American Dictionary:
http://www.collinsdictionary.com/dictionary/american

Notes:

1. Words in black colour indicate foils that have matching vowel to the test item.

2. Words in red colour indicate foils that have either matching initial or final consonants to the test item.

3. Words in green colour indicate foils that bear no relation to the test item.

4. Words that are highlighted with yellow line are the example of test items that their foils differ between lists.

List A1

No.	Test items	Foil items
1	dog	ball, frog, car
2	purse	shirt, girl, bird
3	milk	fish, witch, sink
4	nose	stove, boat, coat
5	hand	ham, man, hair
6	watch	frog, witch, fish
7	foot	fruit, shoe, book
8	smile	knife, light, slide
9	truck	cup, duck, tongue
10	school	shoe, food, spoon
11	door	dog, boy, fork
12	house	mouth, horse, juice
13	bike	slide, knife, light
14	train	cake, grape, tree
15	teeth	tree, meat, foot
16	snake	cake, grape, sink
17	head	bread, hand, stairs
18	bus	duck, dress, book
19	clock	clown, truck, car
20	shoe	school, spoon, juice
21	hair	bear, chair, head
22	tongue	gun, truck, cup
23	witch	fish, sink, watch
24	meat	witch, milk, teeth
25	gum	duck, tongue, gun
26	food	foot, book, shoe
27	soap	coat, boat, stove
28	bird	girl, shirt, purse
29	sink	witch, milk, fish
30	ball	dog, boy, frog
31	tree	teeth, meat, train
32	boat	coat, book, boy
33	duck	truck, dog, book
34	dress	bread, bus, stairs

No.	Test items	Foil items
35	ham	man, pants, hand
36	horse	boy, fork, door
37	light	bike, slide, knife
38	frog	dog, clock,fruit
39	shirt	bird, purse, shoe
40	spoon	School, shoe, food
41	juice	fruit, shoe, school
42	bear	hair, chair, bread
43	girl	purse, bird, shirt
44	cake	snake, grape, car
45	comb	coat, stove, clown
46	coat	comb, boat, nose
47	mouth	clown, house, man
48	man	ham, hand, pants
49	cup	tongue, duck, truck
50	gun	gum, tongue, clown

List A2

No.	Test items	Foil items
1	frog	dog, ball ,car
2	bird	shirt, girl, purse
3	milk	witch, sink, fish
4	boat	nose, stove, coat
5	man	hand, hair, ham
6	watch	frog, witch, fish
7	food	shoe, book foot
8	smile	knife, slide, light
9	duck	tongue, cup, truck
10	spoon	school, shoe, food
11	door	boy, dog, fork
12	mouth	house, horse, juice
13	light	slide, knife, bike
14	train	grape, snake, sink
15	tree	meat, teeth, food
16	cake	snake, grape, sink
17	head	stairs, bread, hand
18	dress	duck, book, bus
19	clock	truck, clown, car
20	juice	spoon, shoe, school
21	bear	hair, chair, head
22	gun	tongue, truck, cup
23	sink	witch, fish, watch
24	meat	milk, witch, teeth
25	tongue	gun, duck, gum
26	foot	book, food, shoe
27	soap	coat, boat, stove
28	girl	bird, shirt, purse
29	witch	fish, sink, milk
30	ball	fork, dog, boy
31	teeth	tree, meat, train
32	coat	boat, boy, book
33	truck	duck, dog, book
34	bus	stairs, bread, dress
35	hand	ham, man, pants

No.	Test items	Foil items
36	horse	boy, door, fork
37	bike	light, knife, slight
38	dog	clock, frog, fruit
39	shirt	bird, purse, shoe
40	school	spoon, shoe, fruit
41	shoe	fruit, juice, school
42	hair	bear, chair, bread
43	purse	bird, shirt, girl
44	snake	cake, grape, car
45	comb	coat, stove, clown
46	nose	comb, coat, boat
47	house	mouth, clown, man
48	ham	hand, pants, man
49	cup	duck, tongue, truck
50	gum	gun, tongue, clown

List B3

No.	Test items	Foil items
1	cake	grape, train, tree
2	comb	boat, coat, stove
3	man	hand, pants, ham
4	soap	coat, boat, stove
5	tree	teeth, meat, train
6	snake	cake, grape, train
7	bus	duck, book, dress
8	juice	school, shoe, fruit
9	gun	tongue, gum, clown
10	cup	tongue, truck, duck
11	meat	teeth, tree, foot
12	bear	chair, hair, head
13	girl	shirt, bird, purse
14	bird	girl, purse, shirt
15	ball	frog, dog, boy
16	food	book, foot, shoe
17	house	mouth, horse, juice
18	spoon	shoe, school, fruit
19	shirt	purse, bird, shoe
20	horse	boy, fork, door
21	light	knife, slide, bike
22	hand	man, ham, hair
23	truck	tongue, duck, cup
24	milk	sink, witch, fish
25	coat	comb, boat, soap
26	frog	dog, ball, boy
27	smile	smile, knife, light
28	mouth	house, horse, juice
29	dress	duck, book, bus
30	witch	fish, sink, milk
31	hair	bear, chair, bread
32	tongue	truck, cup, gun
33	nose	coat, stove, boat
34	gum	tongue, gun, trcuk

No.	Test items	Foil items
35	ham	hand, man, pants
36	shoe	school, spoon, juice
37	train	cake, grape, tree
38	teeth	tree, meat, foot
39	purse	shirt, bird, girl
40	sink	fish, witch, watch
41	school	shoe, fruit, spoon
42	bike	slide, light, smile
43	head	bread, hand, stairs
44	foot	fruit, shoe, book
45	duck	cup, truck, dog
46	door	fork, horse, boy
47	boat	coat, soap, stove
48	watch	fish, witch, frog
49	clock	clown, car, truck
50	dog	frog, clock, fruit

List B4

No.	Test items	Foil items
1	train	grape, cake, tree
2	boat	coat, stove, comb
3	ham	hand, pants, man
4	coat	soap, boat, stove
5	teeth	meat, tree, train
6	snake	cake, grape, train
7	dress	bus, duck, book
8	shoe	school, juice, fruit
9	gum	tongue, gun, clown
10	duck	tongue, truck, cup
11	tree	meat, teeth, foot
12	hair	chair, bear, head
13	girl,	shirt, bird, purse
14	shirt	girl, bird, purse
15	dog	ball, frog, boy
16	foot	book, food, shoe
17	mouth	house, horse, juice
18	school	shoe, spoon, food
19	purse	bird, shirt, shoe
20	door	horse, fork, boy
21	bike	knife, light, slide
22	man	hand, ham, hair
23	cup	tongue, duck, cup
24	sink	milk, witch, fish
25	comb	boat, soap, coat
26	ball	dog, frog, boy
27	light	smile, knife, slide
28	house	mouth, horse, juice
29	bus	duck, book, dress
30	milk	fish, witch, sink
31	bear	hair, chair, bread
32	gun	truck, cup, tongue
33	nose	coat, stove, boat
34	tongue	gum, gun, truck

No.	Test items	Foil items
35	hand	ham, man, pants
36	juice	school, spoon, shoe
37	cake	grape, train, tree
38	meat	teeth, tree, foot
39	bird	shirt, purse, girl
40	witch	fish,witch, watch
41	spoon	shoe, food, school
42	smile	slide, bike, light
43	head	bread, hand, stairs
44	food	foot, shoe, book
45	truck	duck, cup, dog
46	horse	fork, door, boy
47	soap	coat, boat, stove
48	witch	fish, frog, watch
49	clock	clown, truck, car
50	frog	dog, fruit, clock

APPENDIX B:
INITIAL TEST FORMS

Initial four test forms from the same 50 words with different randomisation

	Susunan Tes A1		Susunan Tes A2		Susunan Tes B3		Susunan Tes B4
	Test Form A1		Test form A2		Test form B3		Test form B4
1	Bantal (pillow)	1	Mata (Eye)	1	Lilin (Candle)	1	Lilin (Candle)
2	Kaki (Foot)	2	Bayi (Baby)	2	Topi (Hat)	2	Roti (Bread)
3	Bulan (Moon)	3	Bunga (Flower)	3	Kursi (Chair)	3	Gunting (Scissors)
4	Mata (Eye)	4	Bantal (Pillow)	4	Badut (Clown)	4	Garpu (Fork)
5	Ayam (Chicken)	5	Katak (Frog)	5	Roti (Bread)	5	Topi (Hat)
6	Badut (Clown)	6	Rambut (Hair)	6	Nasi (Rice)	6	Anjing (Dog)
7	Sandal (Slippers)	7	Nanas (Pine Apple)	7	Ember (Bucket)	7	Bebek (Duck)
8	Gajah (Elephant)	8	Gajah (Elephant)	8	Bola (Ball)	8	Coklat (Chocolate)
9	Piring (Plate)	9	Lilin (Candle)	9	Anak (Child)	9	Katak (Frog)
10	Buku (Book)	10	Burung (Bird)	10	Bantal (Pillow)	10	Nanas (Pine Apple)
11	Nanas (Pine Apple)	11	Sandal (Slippers)	11	Rambut (Hair)	11	Badut (Clown)
12	Roti (Bread)	12	Roti (Bread)	12	Ayam (Chicken)	12	Anak (Child)
13	Semut (Ant)	13	Semut (Ant)	13	Nanas (Pine Apple)	13	Mata (Eye)
14	Kera (Monkey)	14	Gelas (Glass)	14	Ikan (Fish)	14	Pisang (Banana)
15	Ember (Bucket)	15	Bebek (Duck)	15	Gelas (Glass)	15	Gelas (Glass)
16	Pisang (Banana)	16	Pisang (Banana)	16	Gajah (Elephant)	16	Tangan (Hand)
17	Jari (Finger)	17	Jari (Finger)	17	Dasi (Tie)	17	Kaki (Foot)
18	Bayi (Baby)	18	Kaki (Foot)	18	Rumah (House)	18	Rumah (House)
19	Buah (Fruit)	19	Buah (Fruit)	19	Bulan (Moon)	19	Bunga (Flower)
20	Bapak (Father)	20	Anak (Child)	20	Lampu (Lamp)	20	Lampu (Lamp)
21	Dasi (Tie)	21	Dasi (Tie)	21	Guling (-)	21	Kucing (Cat)
22	Bunga (Flower)	22	Rumah (House)	22	Telur (Egg)	22	Semut (Ant)
23	Anak (Child)	23	Ayam (Chicken)	23	Bapak (Father)	23	Ayam (Chicken)
24	Katak (Frog)	24	Bapak (Father)	24	Bebek (Duck)	24	Ember (Bucket)
25	Rambut (Hair)	25	Garpu (Fork)	25	Sapi (Cow)	25	Sapi (Cow)
26	Payung (Umbrella)	26	Gayung (Bath Scoop)	26	Bayi (Baby)	26	Bayi (Baby)
27	Lilin (Candle)	27	Piring (Plate)	27	Pisang (Banan)	27	Ikan (Fish)
28	Burung (Bird)	28	Buku (Book)	28	Piring (Plate)	28	Piring (Plate)
29	Topi (Hat)	29	Topi (Hat)	29	Payung (umbrella)	29	Gayung (Bath Scoop)
30	Gelas (Glass)	30	Kera (Monkey)	30	Bunga (Flower)	30	Buah (Fruit)
31	Telur (Egg)	31	Telur (Egg)	31	Tikus (Mice)	31	Hidung (Nose)
32	Bebek (Duck)	32	Ember (Bucket)	32	Mata (Eye)	32	Bantal (Pillow)
33	Ikan (Fish)	33	Ikan (Fish)	33	Tangan (Hand)	33	Gajah (Elephant)
34	Kursi (Chair)	34	Kursi (Chair)	34	Gunting (Scissors)	34	Kursi (Chair)
35	Anjing (Dog)	35	Kambing (Goat)	35	Coklat (Chocolate)	35	Bola (Ball)
36	Lampu (Lamp)	36	Badut (Clown)	36	Buku (Book)	36	Buku (Book)
37	Tangan (Hand)	37	Tangan (Hand)	37	Anjing (Dog)	37	Kambing (Goat)
38	Guling (-)	38	Gunting (Scissors)	38	Buah (Fruit)	38	Bulan (Moon)
39	Sapi (Cow)	39	Anjing (Dog)	39	Burung (Bird)	39	Burung (Bird)
40	Gayung (Bath Scoop)	40	Payung (Umbrella)	40	Kera (Monkey)	40	Kera (Monkey)
41	Garpu (Fork)	41	Lampu (Lamp)	41	Garpu (Fork)	41	Rambut (Hair)
42	Kucing (Cat)	42	Guling (-)	42	Kucing (Cat)	42	Guling (-)
43	Coklat (Chocolate)	43	Bola (Ball)	43	Katak (Frog)	43	Bapak (Father)
44	Tikus (Mice)	44	Hidung (Nose)	44	Gayung (Bath Scoop)	44	Payung (Umbrella)
45	Bola (Ball)	45	Coklat (Chocolate)	45	Jari (Finger)	45	Dasi (Tie)
46	Kambing (Goat)	46	Nasi (Rice)	46	Kaki (Foot)	46	Jari (Finger)
47	Hidung (Nose)	47	Tikus (Mice)	47	Semut (Ant)	47	Telur (Egg)
48	Gunting (Scissors)	48	Kucing (Cat)	48	Hidung (Nose)	48	Tikus (Mice)
49	Rumah (House)	49	Bulan (Moon)	49	Sandal (Slippers)	49	Sandal (Slippers)
50	Nasi (Rice)	50	Sapi (Cow)	50	Kambing (Goat)	50	Nasi (Rice)

APPENDIX C:
266 CONCRETE WORD LIST BASED ON FREQUENCY ORDER

49 rumah	9 batu	5 gunting	3 dahan	2 lantai	1 hewan
44 ikan	9 daun	5 gunung	3 danau	2 lidah	1 injil
33 kera	9 gajah	5 ibu	3 duyung	2 monyet	1 intan
36 kaki	9 guling	5 jagung	3 itik	2 motor	1 kendi
27 buku	9 orang	5 jahe	3 jeruk	2 nangka	1 kertas
27 tangan	9 perut	5 kambing	3 kakek	2 obat	1 keset
26 bayi	9 piring	5 kapal	3 keong	2 otak	1 ketel
26 kucing	9 tikus	5 kerbau	3 lalat	2 padi	1 kijang
25 bola	8 anjing	5 kompor	3 leher	2 paha	1 kodok
25 pisang	8 bintang	5 kuku	3 lutut	2 paku	1 koran
24 burung	8 dasi	5 muka	3 naga	2 panah	1 korek
24 loli	8 mangga	5 obeng	3 nenek	2 pipa	1 kunci
23 anak	8 sendok	5 panci	3 penyu	2 poci	1 laso
21 bunga	8 wajan	5 pisau	3 permen	2 pusar	1 lembu
21 payung	7 air	5 ubi	3 punggung	2 quran	1 limau
20 pensil	7 api	5 yoyo	3 rumput	2 ranting	1 lonceng
20 pohon	7 bantal	4 bajing	3 sapu	2 robot	1 manggis
20 baju	7 garpu	4 bebek	3 sawah	2 roda	1 mangkok
18 mata	7 kayu	4 cecak	3 sayur	2 rusa	1 martil
17 buah	7 lampu	4 foto	3 tumit	2 sayap	1 mawar
17 gigi	7 langit	4 handuk	3 ulat	2 serbet	1 merak
17 mulut	7 nyamuk	4 jambu	3 wayang	2 tangga	1 mesjid
16 ayam	7 pintu	4 kado	2 atap	2 tiang	1 panda
14 kancil	7 wortel	4 obor	2 bakso	2 unta	1 pena
13 jari	7 tali	4 pagar	2 bangau	2 vandel	1 pesut
13 meja	7 bapak	4 palu	2 cangkul	2 weker	1 pompa
13 sapi	6 ember	4 pantai	2 capung	1 anggur	1 rantai
13 tubuh	6 gayung	4 raja	2 cincin	1 angklung	1 ratu
12 tupai	6 lebah	4 salak	2 dadu	1 asbak	1 roket
11 badut	6 musang	4 sangkar	2 dahi	1 awan	1 rokok
11 gelas	6 odol	4 sarang	2 gitar	1 cacing	1 rubah
11 hidung	6 pantat	4 siku	2 golok	1 candi	1 sawo
11 mobil	6 papan	4 sisir	2 guci	1 ceret	1 sekop
11 nasi	6 sabun	4 susu	2 hiu	1 cobek	1 senter
11 roti	6 sandal	4 telur	2 jamur	1 congklak	1 singa
11 rambut	6 semut	4 udang	2 jarum	1 cula	1 siput
10 ayah	6 sikat	4 zebra	2 kail	1 delman	1 sumur
10 coklat	6 tulang	3 babi	2 kapur	1 domba	1 surat
10 katak	6 ular	3 bambu	2 kasur	1 dompet	1 tomat
10 kursi	6 tulang	3 batang	2 kaus	1 dongkrak	1 tugu
10 lilin	5 biji	3 benang	2 kompor	1 kuku	1 wajah
10 nanas	5 cabai	3 bibir	2 koper	1 fliper	
10 topi	5 dada	3 botol	2 kuda	1 gasing	
10 badan	5 ekor	3 cermin	2 kumbang	1 gaun	
9 apel	5 elang	3 daging	2 labu	1 gua	

APPENDIX D:
INDO-SPRITT FOILS

List A1

No.	Test items	Number of effective foils	Foil items
1	Bantal	3	Sandal, nanas, mata
2	Kaki	3	Babi, bayi, nasi
3	Bulan	3	Buah, rumah, bunga
4	Mata	3	Bantal, sandal, nanas
5	Ayam	3	Anak, bapak, kapal
6	Badut	3	Sapu, rambut, anggur
7	Sandal	3	Nanas, bantal, anak
8	Gajah	3	Nanas, tangan, kapal
9	Piring	3	Lilin, gigi, sisir
10	Buku	3	Burung, gunung, mulut
11	Nanas	3	Bantal, sandal, anak
12	Roti	2	Topi, poci, mobil
13	Semut	2	Telur, penyu, jeruk
14	Cecak	1	Kera, elang, gelas
15	Ember	2	Nenek, bebek, ceret
16	Pisang	3	Ikan, kijang, lidah
17	Jari	3	Nasi, babi, sapi
18	Bayi	3	Babi, dasi, kaki
19	Buah	3	Bunga, bulan, rumah
20	Bapak	3	Anak, ayam, gajah
21	Dasi	3	Nasi, jari, kaki
22	Bunga	3	Kuda, buah, rumah
23	Anak	3	Ayam, gajah, tnagan
24	Kapal	3	Bapak, nanas, mata
25	Rambut	3	Lampu, badut, garpu
26	Payung	3	Badut, gayung, jagung
27	Lilin	3	Piring, gigi, sisir
28	Burung	3	Gunung, mulut, buku
29	Topi	2	Mobil, roti, poci
30	Gelas	1	Cecak, kera, elang
31	Telur	2	Penyu, semut jeruk

No.	Test items	Number of effective foils	Foil items
32	Bebek	2	Ceret, nenek. ember
33	Ikan	3	Lidah, kijang, pisang
34	Kursi	3	Kucing, kursi, gunting
35	Anjing	3	Nasi, jari, kambing
36	Lampu	3	Garpu, badut, rambut
37	Tangan	3	Anak, ayam, gajah
38	Guling	3	Gunting, kucing, kunci
39	Sapi	3	Kambing, jari, anjing
40	Gayung	3	Paying, jagung, rambut
41	Garpu	3	Lampu, sapu, gayung
42	Kucing	3	Kunci, kursi, guling
43	Coklat	3	Domba, bola, tomat
44	Tikus	3	Ibu, pintu, hidung
45	Bola	3	Coklat, tomat domba
46	Kambing	3	Nasi, anjing, jari
47	Hidung	3	Pintu, ibu, tikus
48	Gunting	3	Guling, kucing, kunci
49	Rumah	3	Ular, bulan, buah
50	Nasi	3	Dasi, sapi, babi

List A2

No.	Test items	Number of effective foils	Foil items
1	Mata	3	Bantal, sandal, nanas
2	Bayi	3	Babi, kaki, nasi
3	Bunga	3	Bulan, rumah, buah
4	Bantal	3	Sandal, mata, nanas
5	Kapal	3	Anak, ayam, bapak
6	Rambut	3	Sapu, badut, anggur
7	Nanas	3	Sandal, bantal, anak
8	Gajah	3	Nanas, tangan, katak
9	Lilin	3	Piring, gigi, sisir
10	Burung	3	Gunung, buku, mulut
11	Sandal	3	Bantal, anak, nanas
12	Roti	2	Topi, poci, mobil
13	Semut	2	Telur, penyu, jeruk
14	Gelas	1	Kera, elang, cecak
15	Bebek	2	Nenek, ember, ceret
16	Pisang	3	Ikan, kijang, lidah
17	Jari	3	Nasi, babi, sapi
18	Kaki	3	Bayi, babi, dasi
19	Buah	3	Bunga, bulan, rumah
20	Anak	3	Bapak, ayam, gajah
21	Dasi	3	Nasi, jari, kaki
22	Rumah	3	Anak, gajah, tangan
23	Ayam	3	Anak, gajah, tangan
24	Bapak	3	Nanas, mata, kapal
25	Garpu	3	Lampu, rambut, badut
26	Gayung	3	Badut, payung, jagung
27	Piring	3	Lilin, gigi, sisir
28	Buku	3	Burung, gunung, mulut
29	Topi	2	Mobil, roti, poci
30	Cecak	1	Kera, gelas, elang
31	Telur	2	Penyu, semut, jeruk
32	Ember	2	Ceret, nenek, bebek
33	Ikan	3	Lidah, kijang, pisang
34	Kursi	3	Kucing, kunci, gunting

No.	Test items	Number of effective foils	Foil items
35	Kambing	3	Nasi, jari, anjing
36	Badut	3	Garpu, lampu, rambut
37	Tangan	3	Anak, ayam, gajah
38	Gunting	3	Guling, kucing, kunci
39	Anjing	3	Kambing, sapi, jari
40	Payung	3	Jagung, gayung, rambut
41	Lampu	3	Garpu, sapu, gayung
42	Guling	3	Kunci, kursi, kucing
43	Bola	3	Coklat, domba, tomat
44	Hidung	3	Ibu, tikus, pintu
45	Coklat	3	Bola, tomat, domba
46	Nasi	3	Anjing, kambing, jari
47	Tikus	3	Hidung, pintu, ibu
48	Kucing	3	Guling, kunci, gunting
49	Bulan	3	Ular, rumah, buah
50	Sapi	3	Dasi, babi, nasi

Note: In Book A
Test items with 1 effective foil are in page 14 and page 30
Test items with 2 effective foils are in page 12, 13, 14, 15, 29, 31, 32 (6 pages)
Test items with 3 effective foils are in the rest of the pages (42 pages)

List B3

No.	Test items	Number of effective foils	Foil items
1	Lilin	3	Sisir, gigi, piring
2	Topi	2	Poci, roti, mobil
3	Kursi	3	Kucing, kunci, gunting
4	Badut	3	Garpu, lampu, rambut
5	Roti	2	Mobil, topi, poci
6	Nasi	3	Anjing, jari, kambing
7	Ember	2	Ceret, nenek, bebek
8	Bola	3	Domba, coklat, tomat
9	Anak	3	Kapal, ayam, bapak
10	Bantal	3	Mata, sandal, nanas
11	Rambut	3	Badut, lampu, sapu
12	Ayam	3	Anak, kapal, tangan
13	Nanas	3	Tangan, mata, anak
14	Ikan	3	Pisang, kijang, lidah
15	Gelas	1	Cecak, elang, kera
16	Gajah	3	Katak, nanas, tangan
17	Dasi	3	Nasi, kaki, jari
18	Rumah	3	Buah, ular, bulan
19	Bulan	3	Ular, kuda, bunga
20	Lampu	3	Sapu, garpu, anggur
21	Guling	3	Gunting, kucing, kunci
22	Telur	2	Jeruk, semut, penyu
23	Bapak	3	Anak, katak, ayam
24	Bebek	2	Nenek, ember, ceret
25	Sapi	3	Nasi, babi. kaki
26	Bayi	3	Dasi, babi, kaki
27	Pisang	3	Ikan, lidah, kijang
28	Piring	3	Gigi, sisir, lilin
29	Payung	3	Gayung, anggur, jagung
30	Bunga	3	Rumah, bunga, kuda
31	Tikus	3	Hidung, ibu, pintu
32	Mata	3	Bantal, sandal, kapal
33	Tangan	3	Anak, gajah, kapal
34	Gunting	3	Guling, kunci, kursi

No.	Test items	Number of effective foils	Foil items
35	Coklat	3	Bola, tomat, domba
36	Buku	3	Gunung, burung, buku
37	Anjing	3	Jari, babi, kambing
38	Buah	3	Bunga, rumah, bulan
39	Burung	3	Gunung, buku, mulut
40	Cecak	1	Gelas, elang, kera
41	Garpu	3	Rambut, badut, lampu
42	Kucing	3	Gunting, guling, kunci
43	Kapal	3	Bapak, anak, gajah
44	Gayung	3	Payung, garpu, jagung
45	Jari	3	Kaki, nasi, dasi
46	Kaki	3	Dasi, nasi, jari
47	Semut	2	Penyu, telur , jeruk
48	Hidung	3	Ibu, tikus, pintu
49	Sandal	3	Tangan, bantal, kapal
50	Kambing	3	Anjing, dasi, nasi

List B4

No.	Test items	Number of effective foils	Foil items
1	Lilin	3	Sisir, gigi, piring
2	Roti	2	Poci, topi, mobil
3	Gunting	3	Kursi, kucing, kursi
4	Garpu	3	Badut, lampu, rambut
5	Topi	2	Mobil, poci, roti
6	Anjing	3	Nasi, jari, kambing
7	Bebek	2	Ceret, nenek, ember
8	Coklat	3	Bola, domba, tomat
9	Kapal	3	Ayam, anak, bapak
10	Nanas	3	Mata, bantal, sandal
11	Badut	3	Lampu, rambut, sapu
12	Anak	3	Ayam, kapal, tangan
13	Mata	3	Tangan, nanas, anak
14	Pisang	3	Kijang, lidah, ikan
15	Gelas	1	Cecak, elang, kera
16	Tangan	3	Kapal, gajah, nanas
17	Kaki	3	Dasi, nasi, jari
18	Rumah	3	Buah, ular, bulan
19	Bunga	3	Ular, kuda,bulan
20	Lampu	3	Sapu, garpu, anggur
21	Kucing	3	Gunting, guling kunci
22	Semut	2	Jeruk, telur, penyu
23	Ayam	3	Bapak, anak, kapal
24	Ember	2	Nenek, bebek, ceret
25	Sapi	3	Nasi, babi,kaki
26	Bayi	3	Dasi, babi, kaki
27	Ikan	3	Lidah, kijang, pisang
28	Piring	3	Gigi, sisir, lilin
29	Gayung	3	Payung, anggur, jagung
30	Buah	3	Rumah, bunga, kuda
31	Hidung	3	Tikus, ibu, pintu
32	Bantal	3	Sandal, kapal, mata
33	Gajah	3	Anak, tangan, kapal
34	Kursi	3	Gunting, guling kunci

No.	Test items	Number of effective foils	Foil items
35	Bola	3	Coklat, tomat, domba
36	Buku	3	Gunung, burung, mulut
37	Kambing	3	Jari, babi, anjing
38	Bulan	3	Buah, bunga, rumah
39	Burung	3	Gunung, buku, mulut
40	Cecak	1	Gelas, elang, kera
41	Rambut	3	Garpu, badut, lampu
42	Guling	3	Kucing, gunting, kunci
43	Bapak	3	Anak, gajah, kapal
44	Payung	3	Garpu, jagung, gayung
45	Dasi	3	Jari, kaki, nasi
46	Jari	3	Dasi, nasi, kaki
47	Telur	2	Semut, penyu, jeruk
48	Tikus	3	Ibu, hidung, pintu
49	Sandal	3	Tangan, bantal, bapak
50	Nasi	3	Anjing, kambing, dasi

Note: In Book B
Test items with 1 effective foil are in page 15 and page 40
Test items with 2 effective foils are n page 2, 5, 7, 15, 22, 24, 47 (6 pages)
Test items with 3 effective foils in the rest of the pages (42 pages)

APPENDIX E:
PICTURE BOOKS A & B

Refer to the following pages for Picture Book A (51 pages) and Picture Book B (51 pages)

INDO-SPRITT

Indonesian – Speech Recognition Threshold Test
(Tes ambang pendengaran dengan tutur kata bahasa Indonesia)

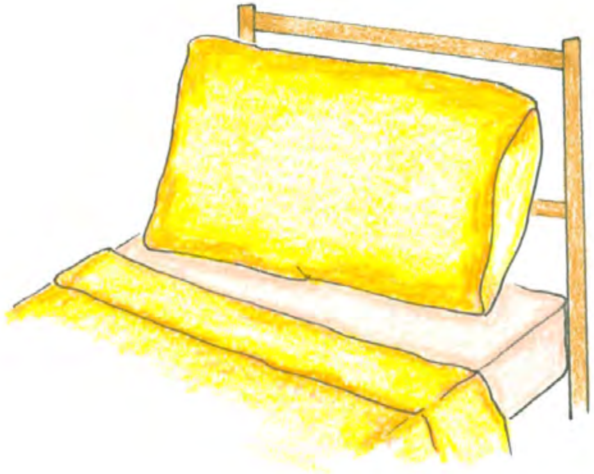
Book A (Test form 1 & 2)
Buku A (Susunan tes 1 & 2)

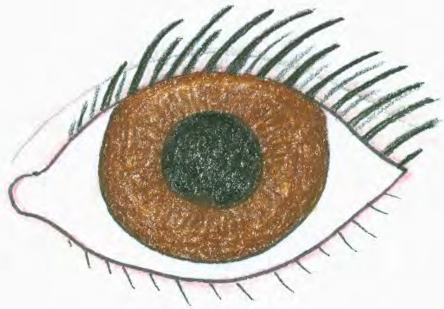
Dahlia E Sartika

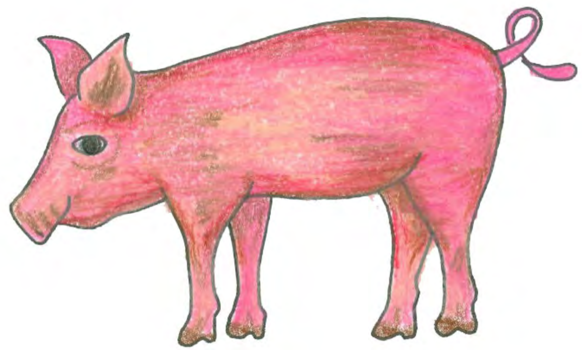
Philip Newall, Mridula Sharma, Robert Mannell, Harvey Dillon

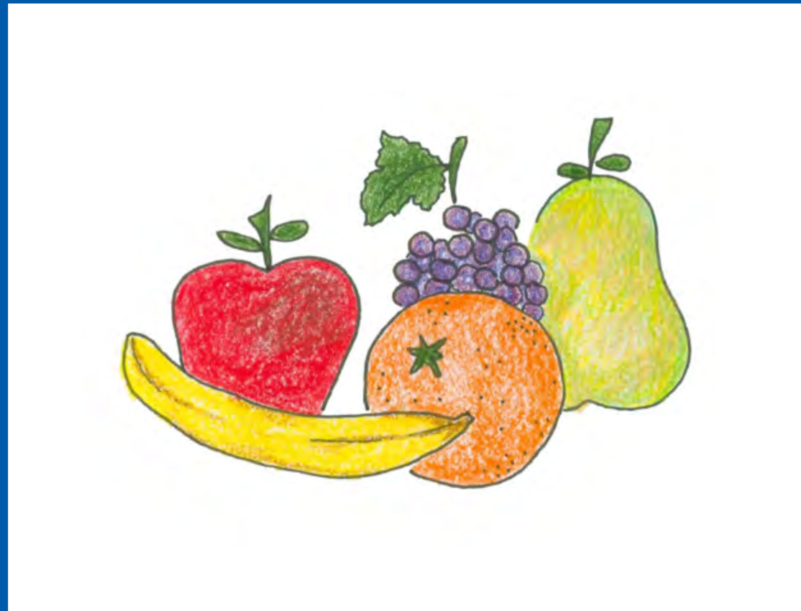
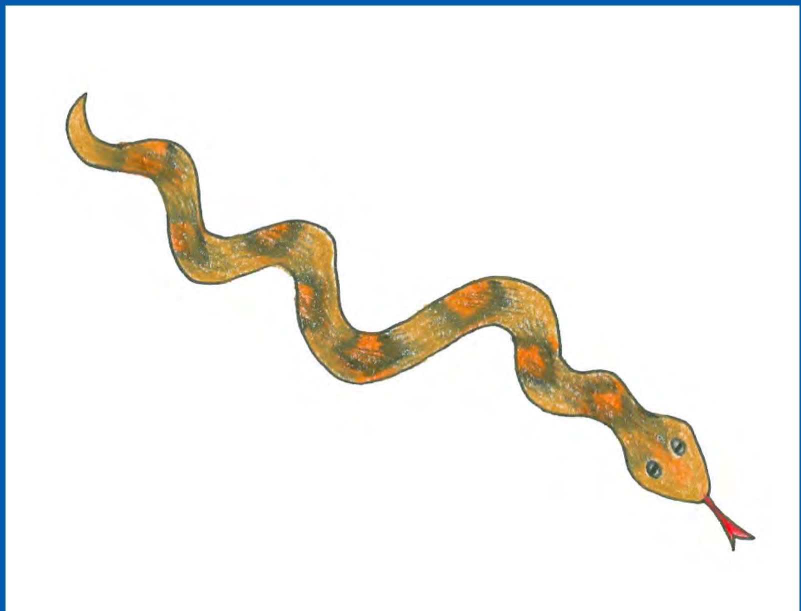
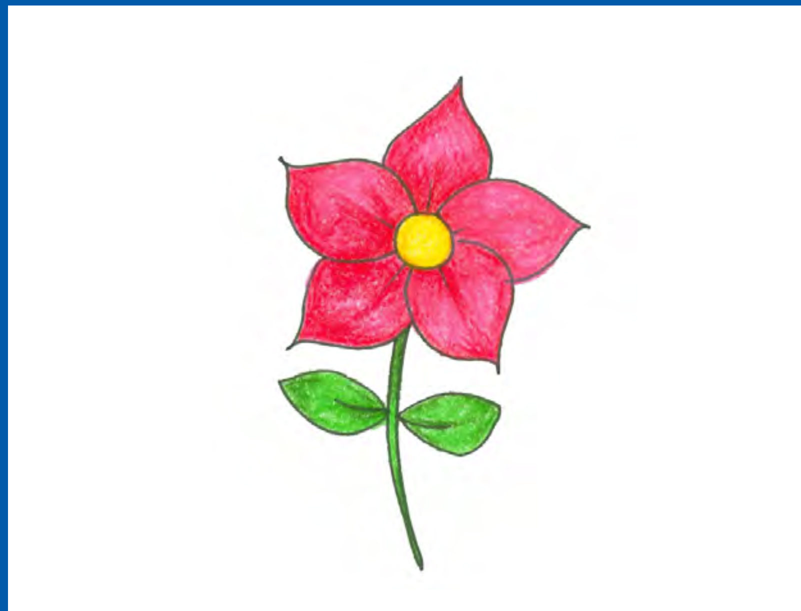
MACQUARIE UNIVERSITY SYDNEY – AUSTRALIA

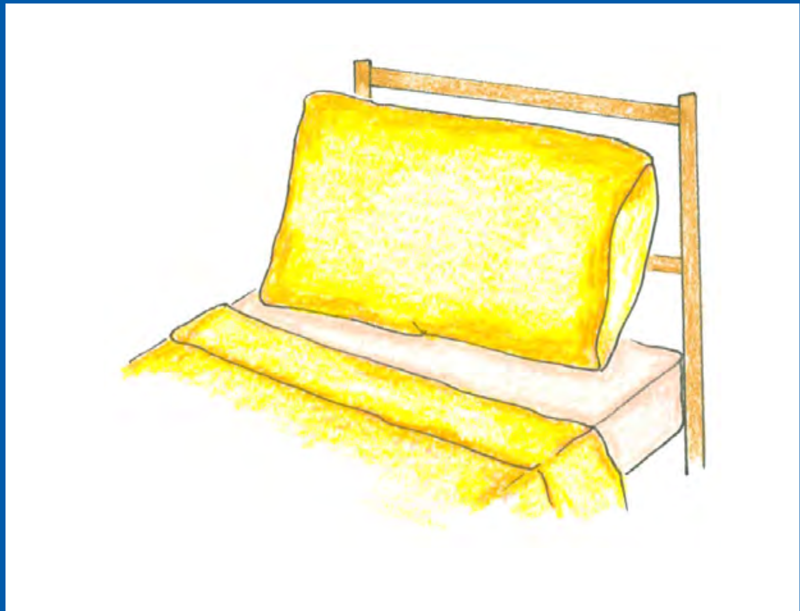
RIDBC Renwick Centre For Research and Professional Education

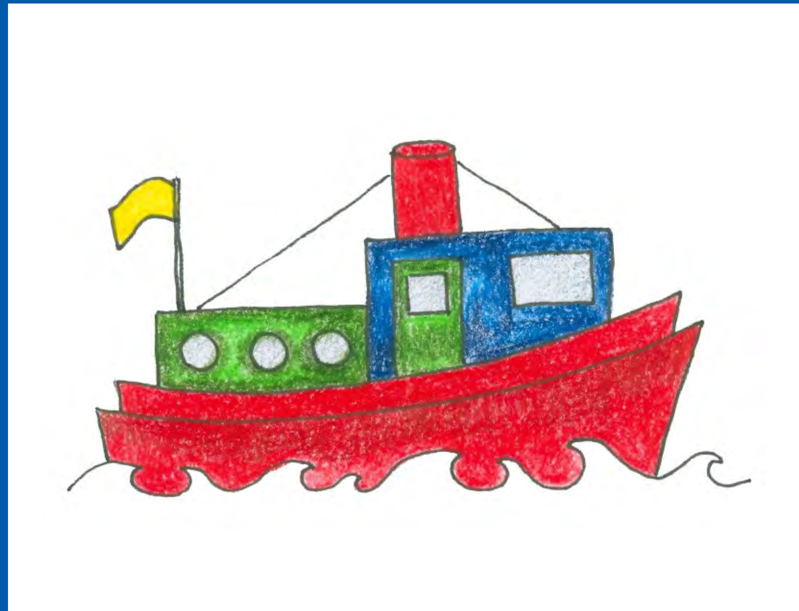


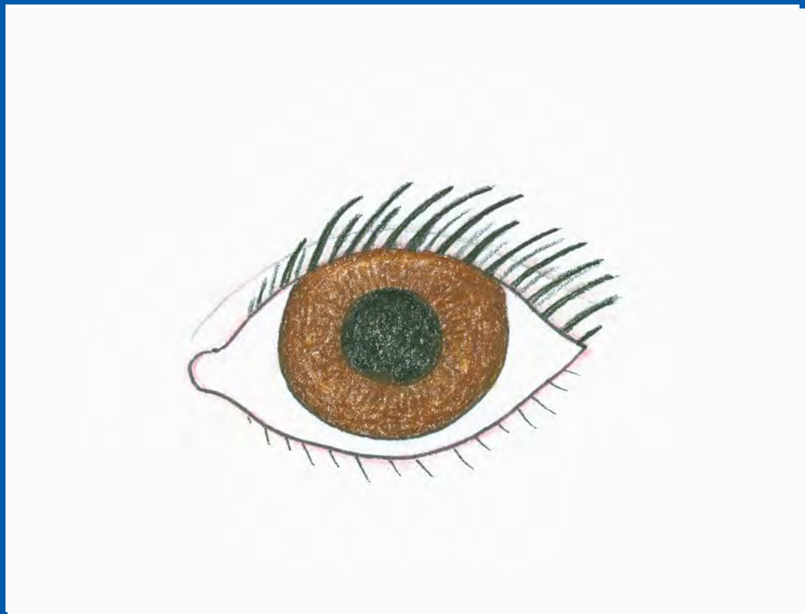



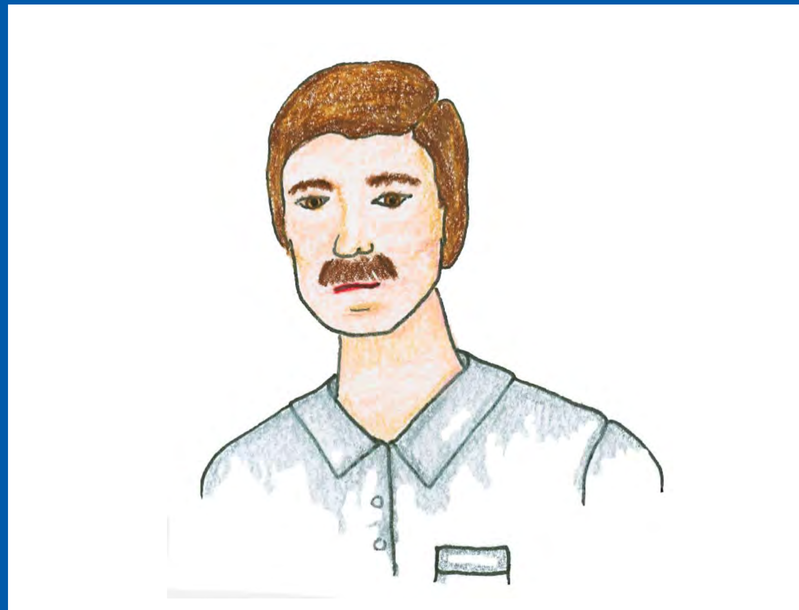


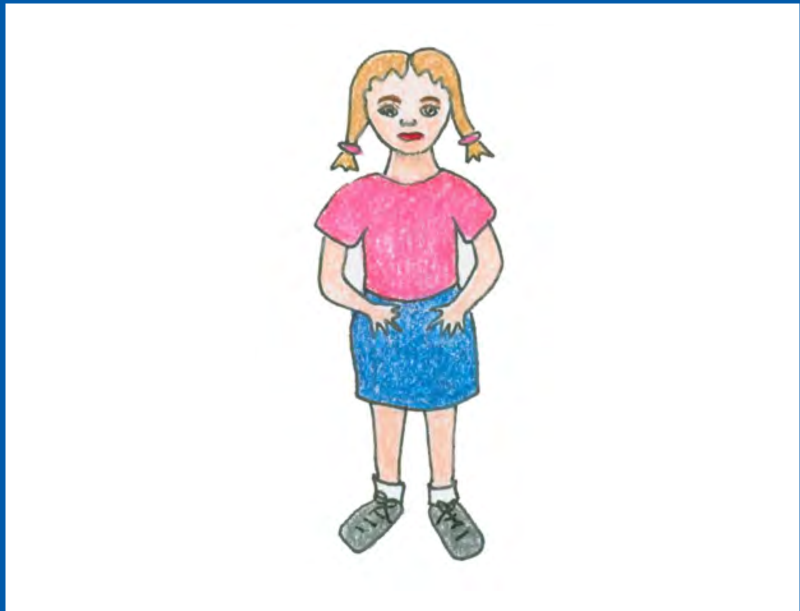
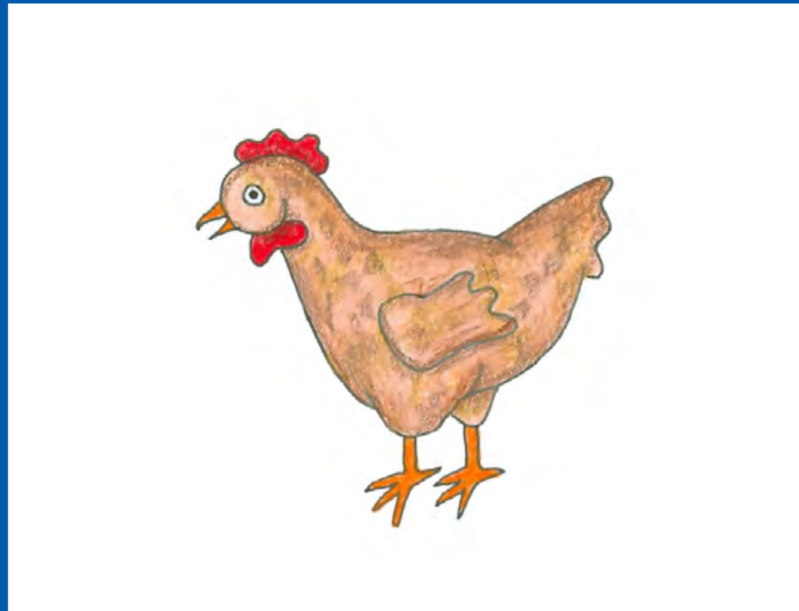
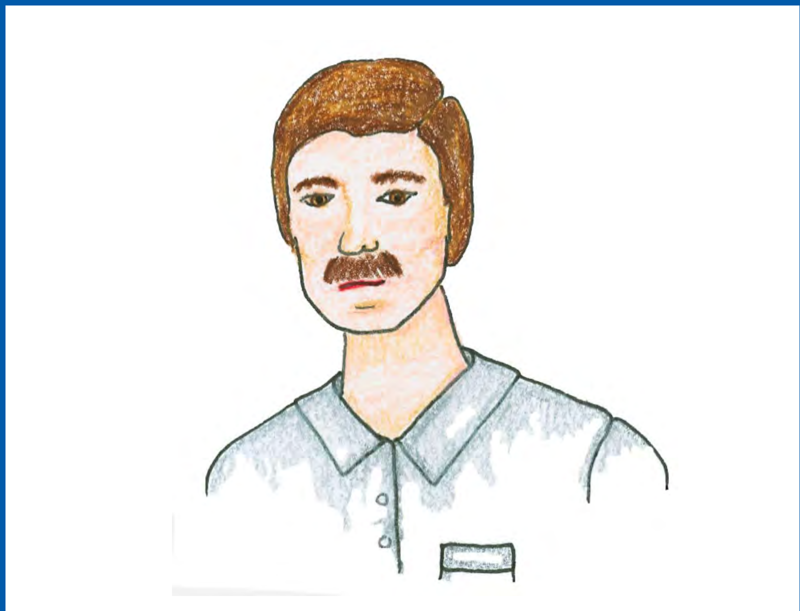
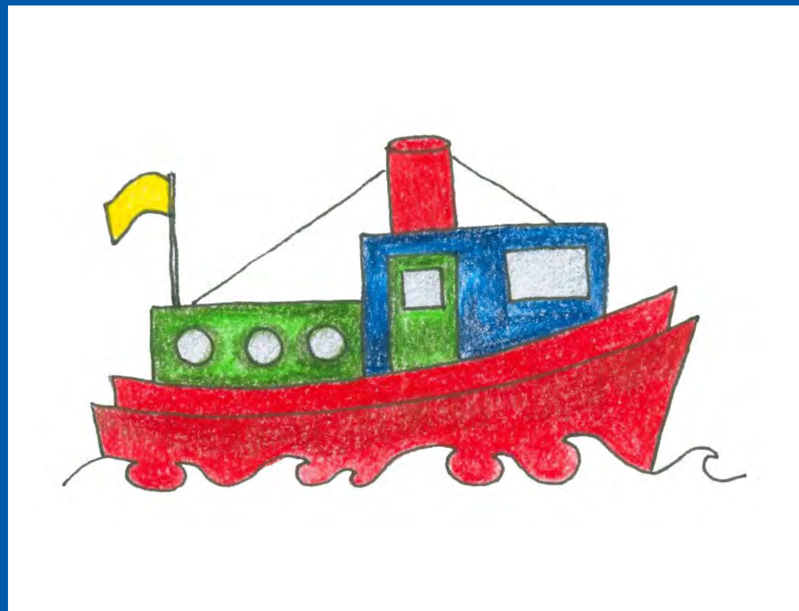





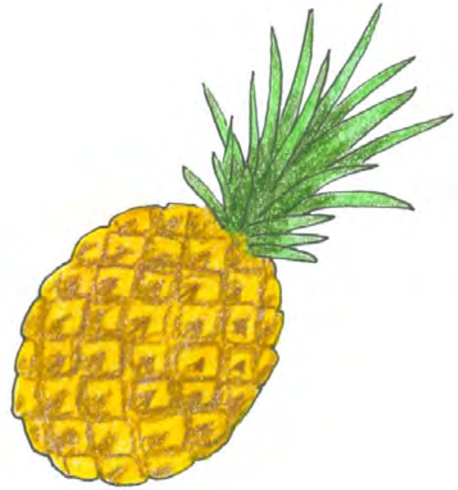
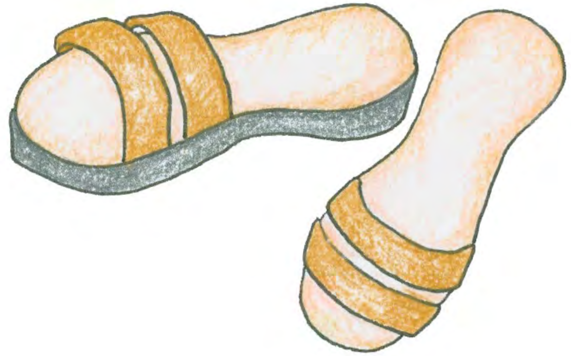
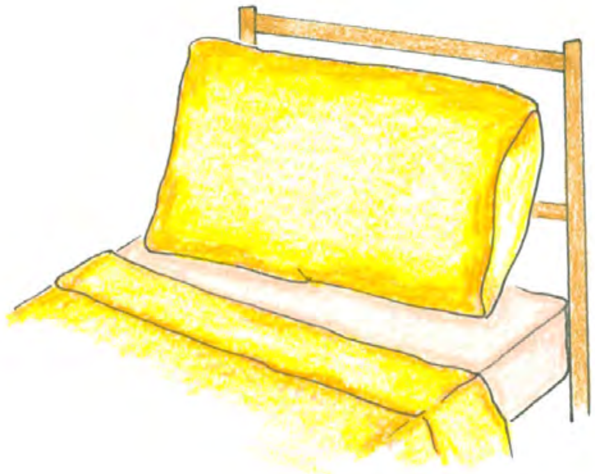


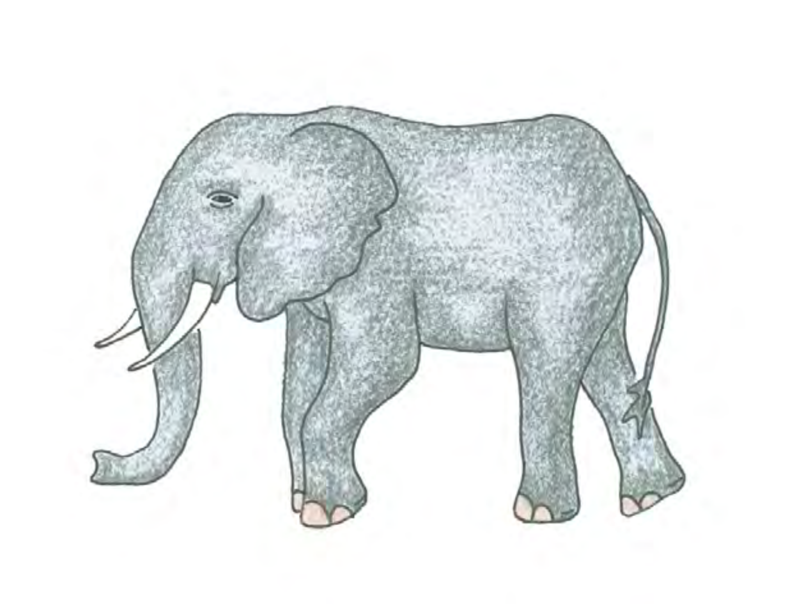
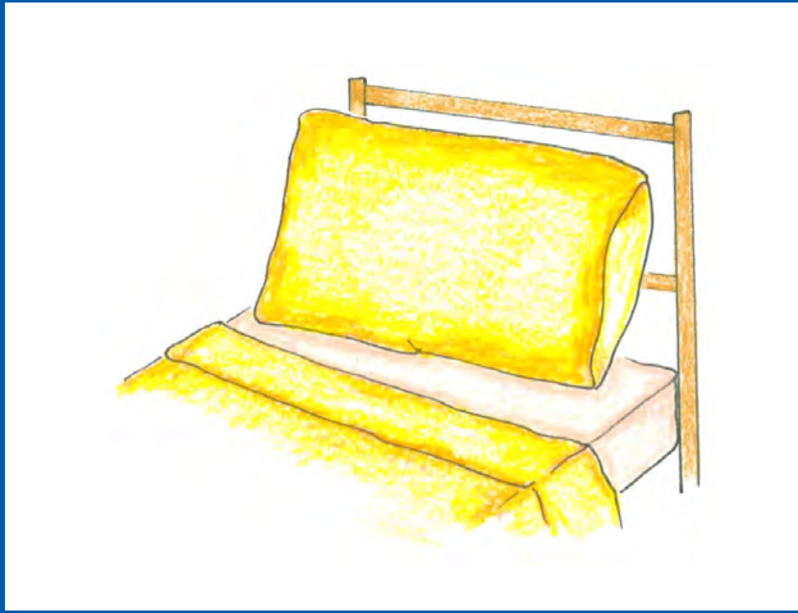


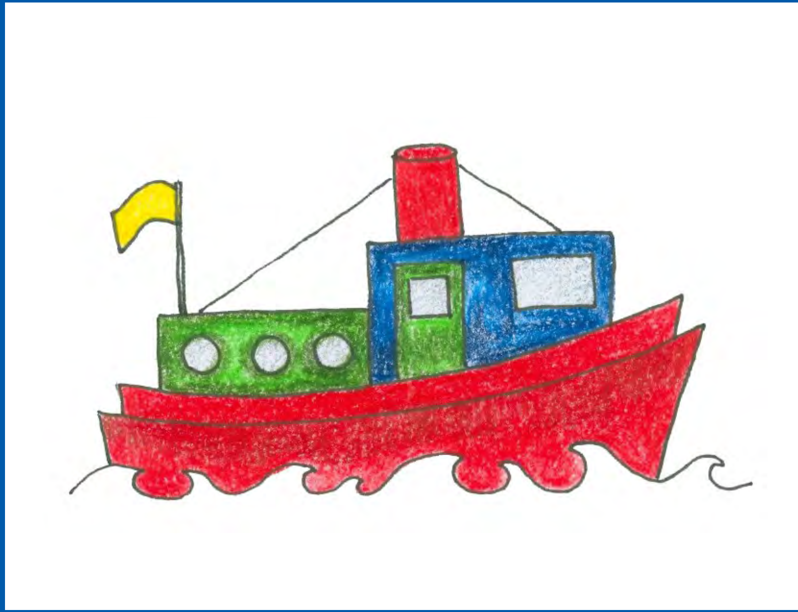


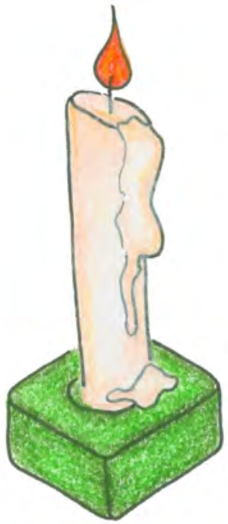


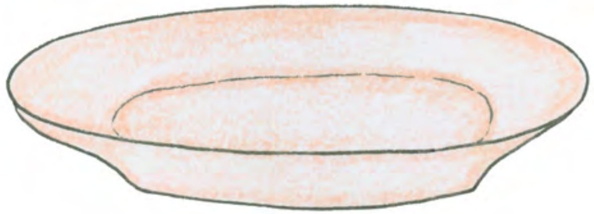





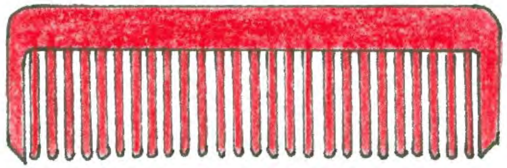




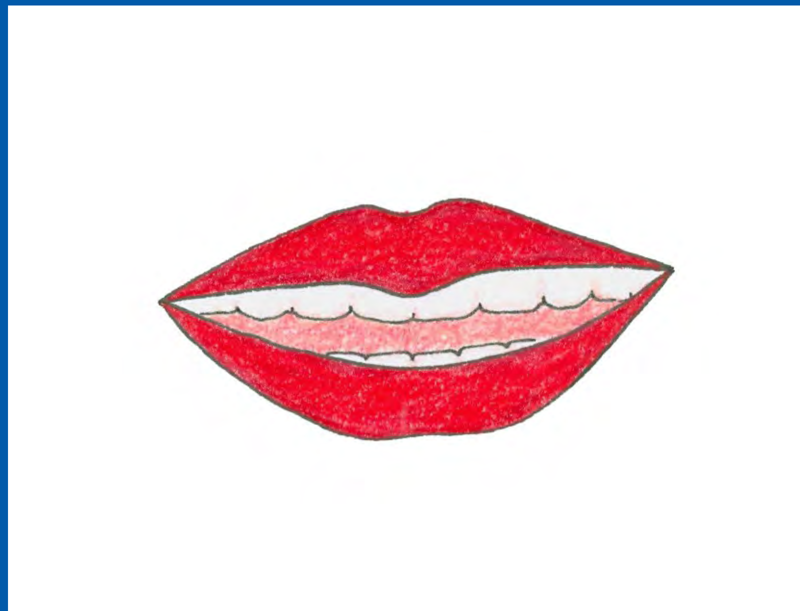




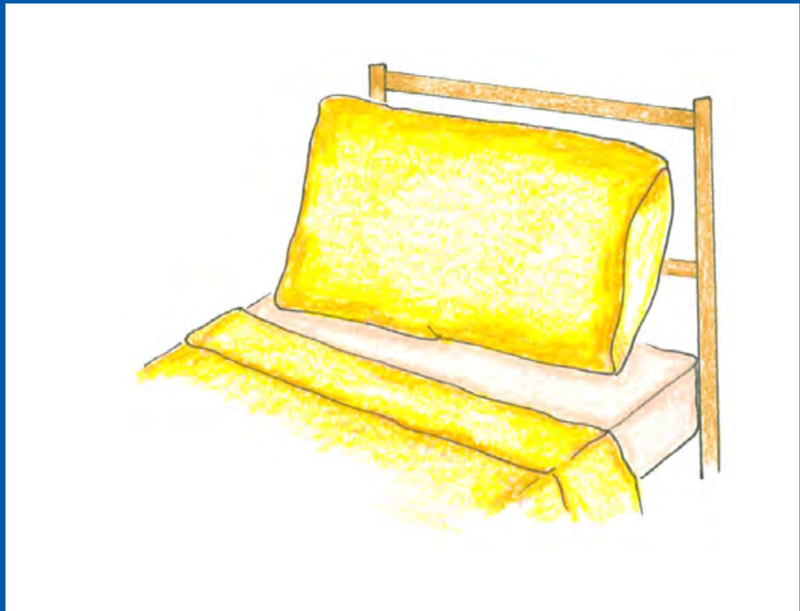


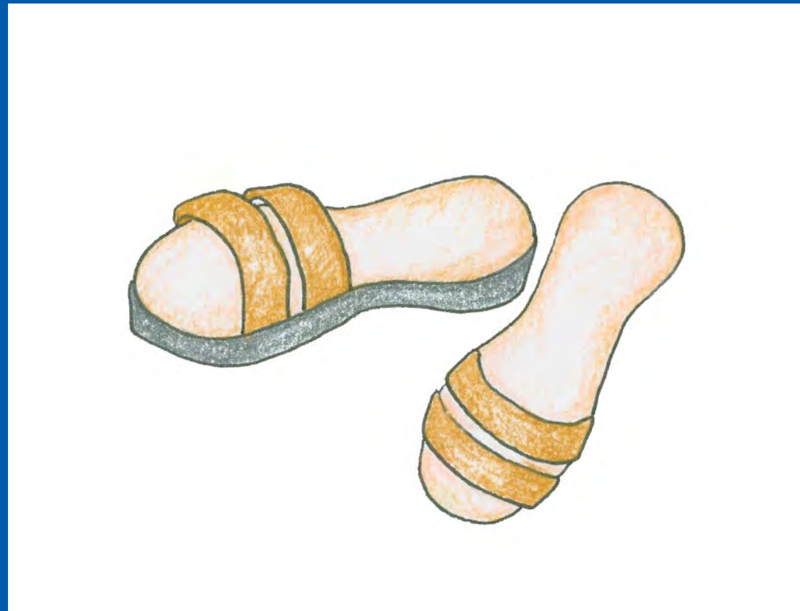


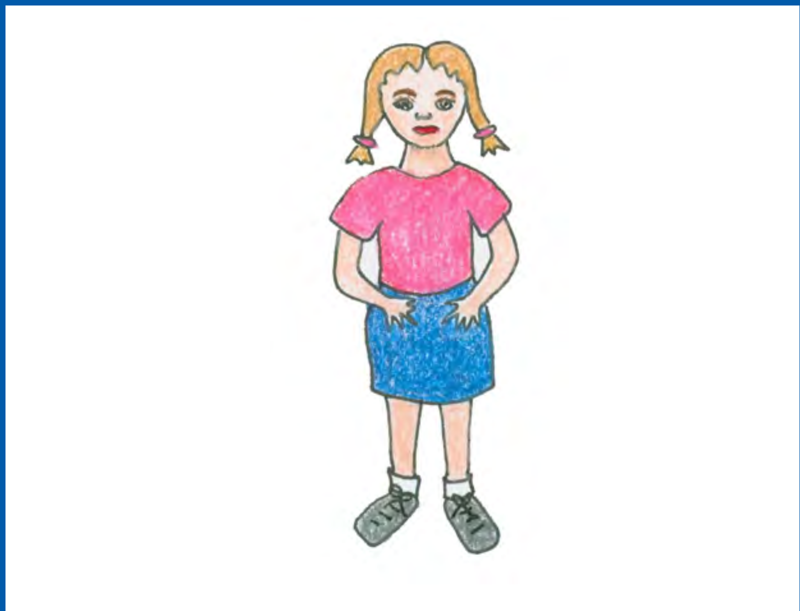


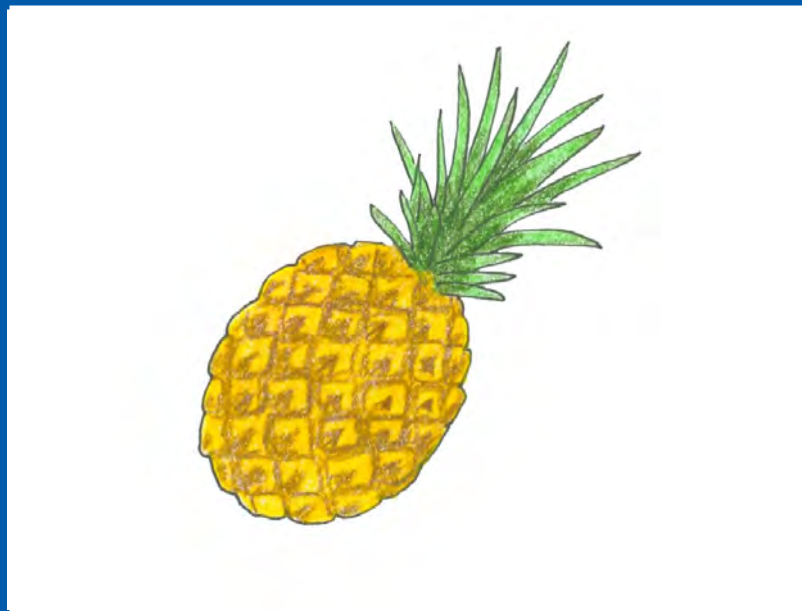


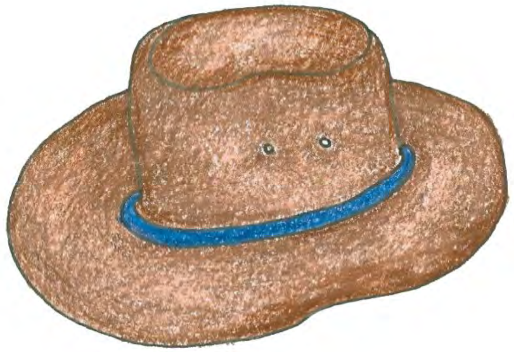


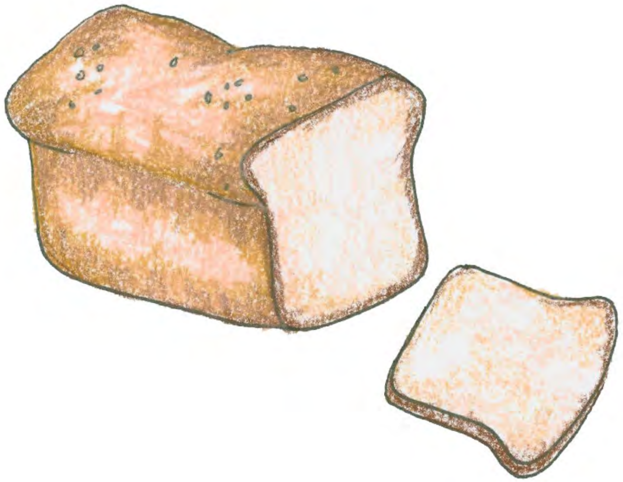


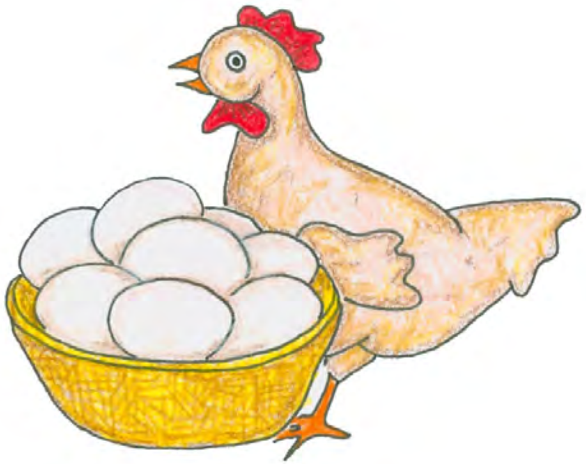
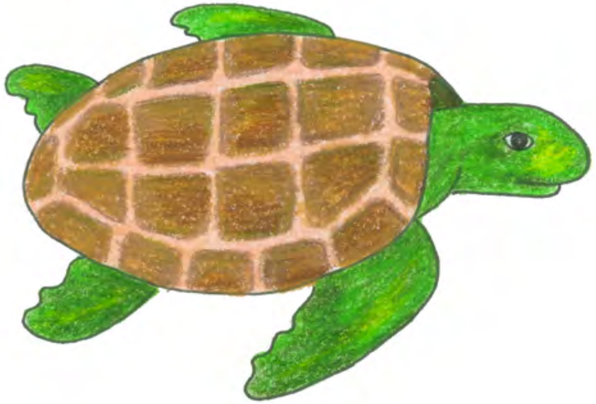


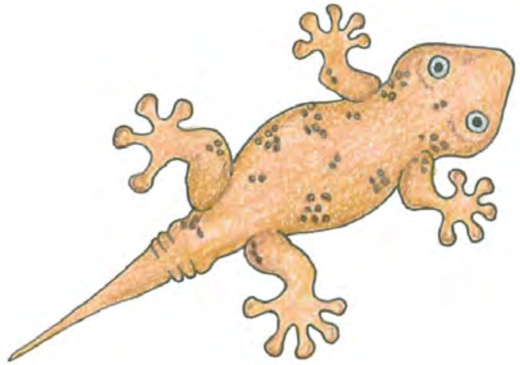


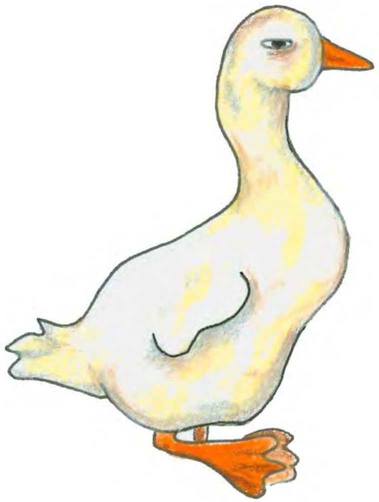


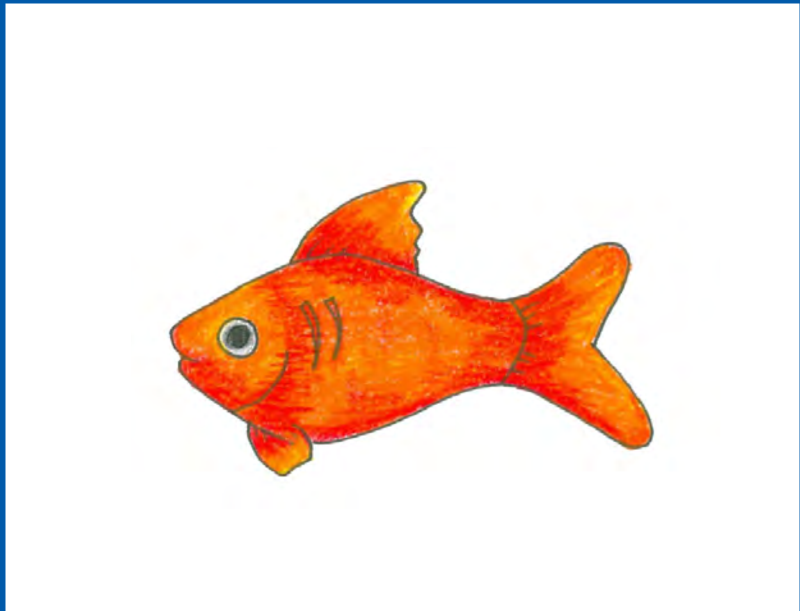
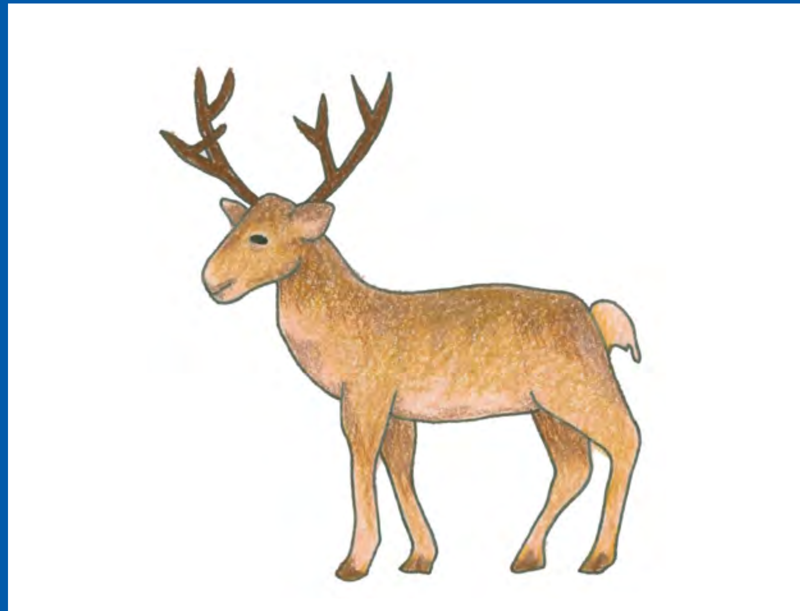
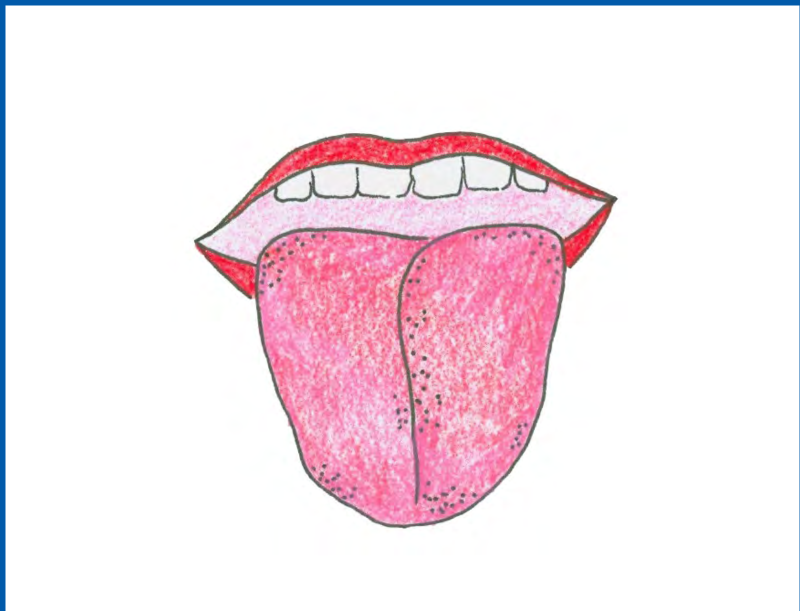
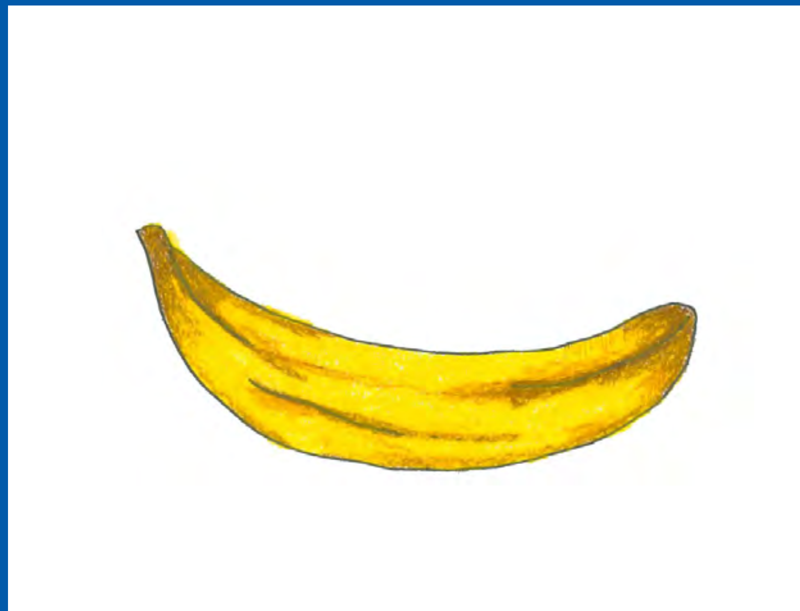


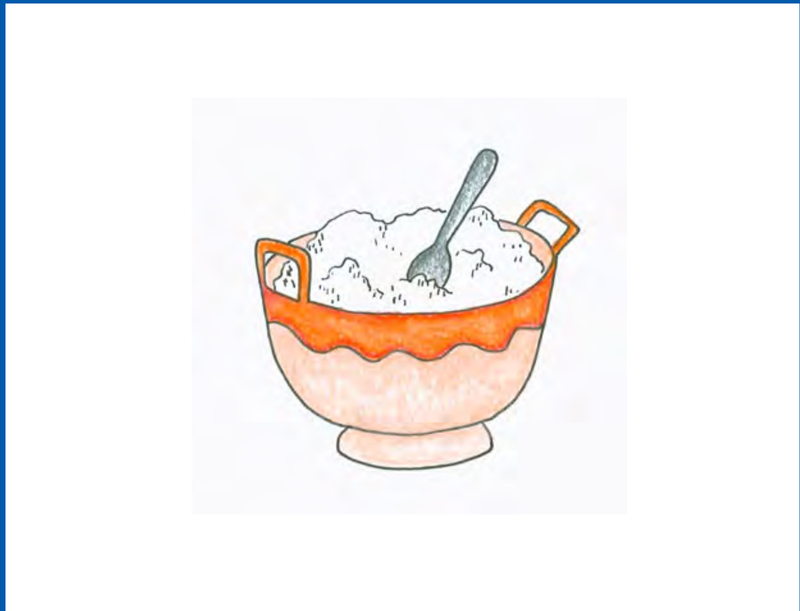


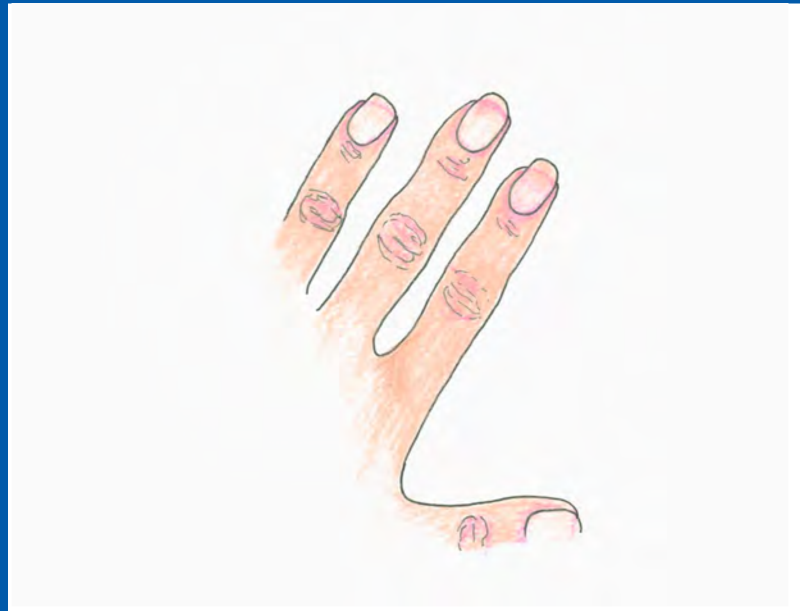




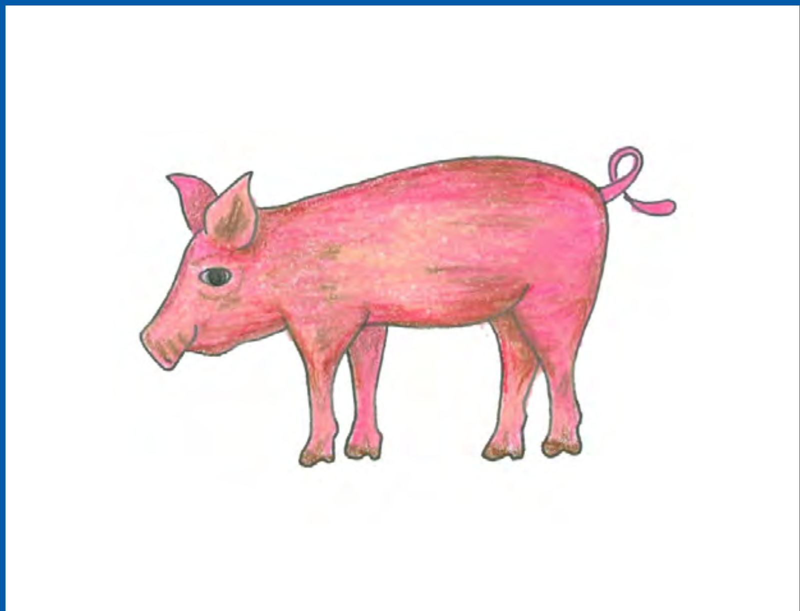


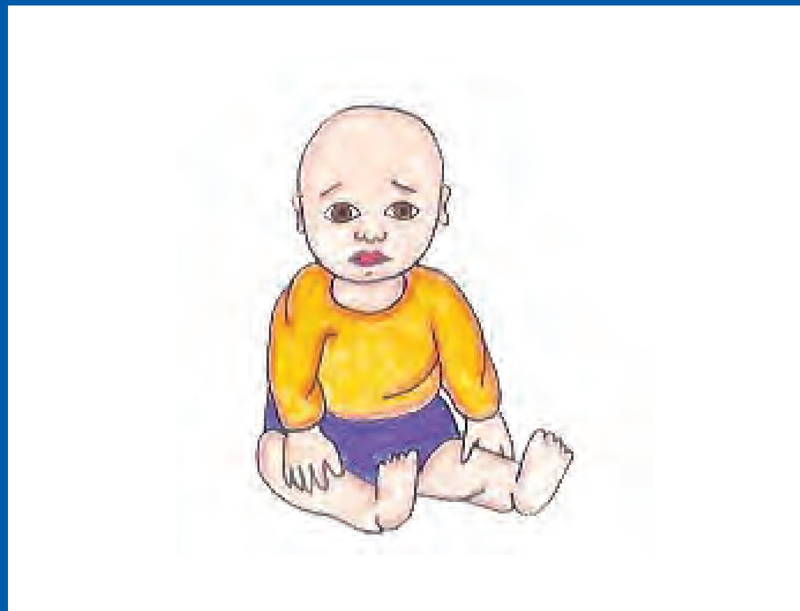


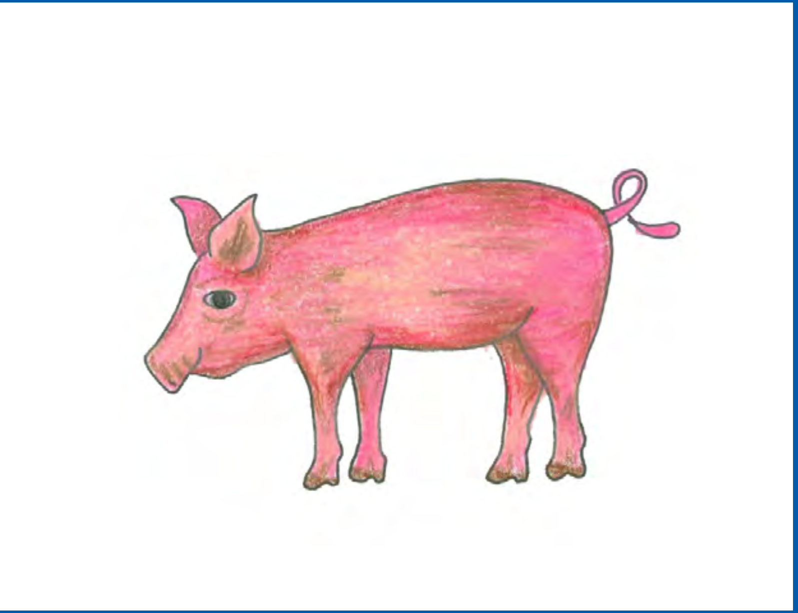









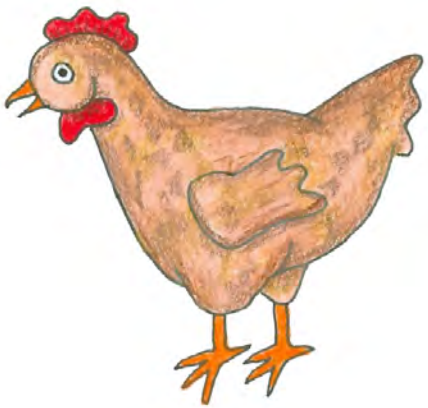
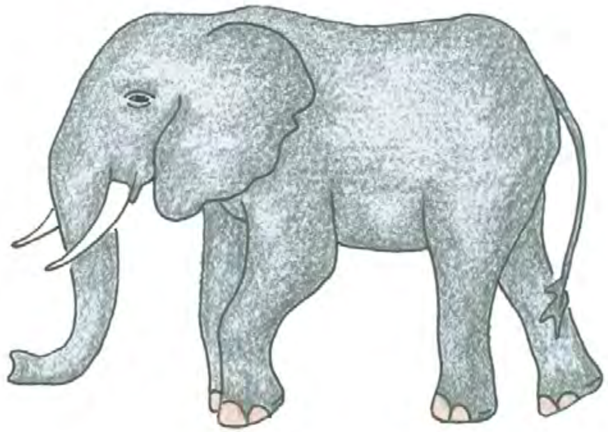


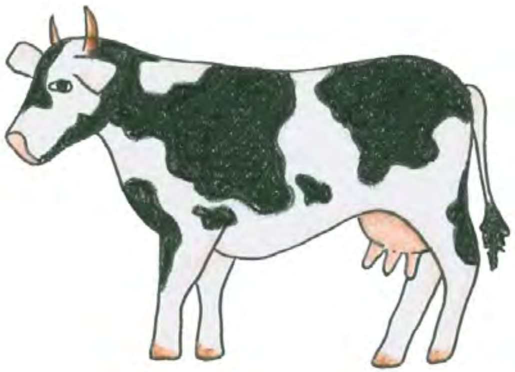


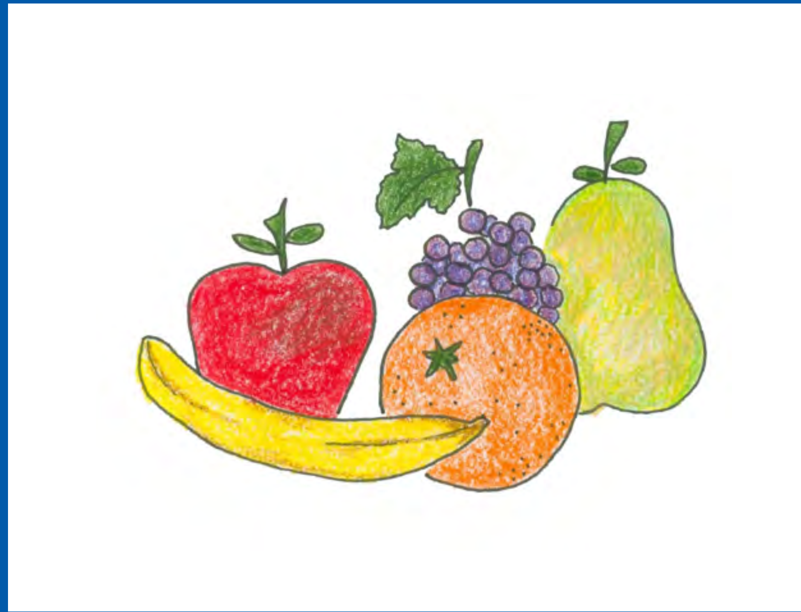


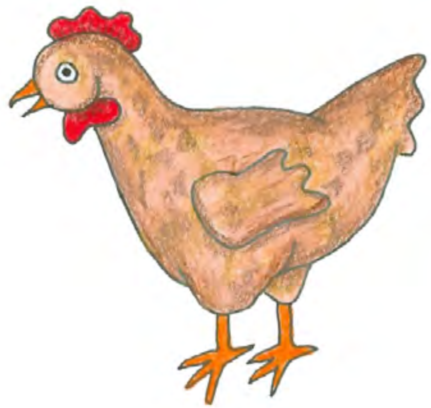


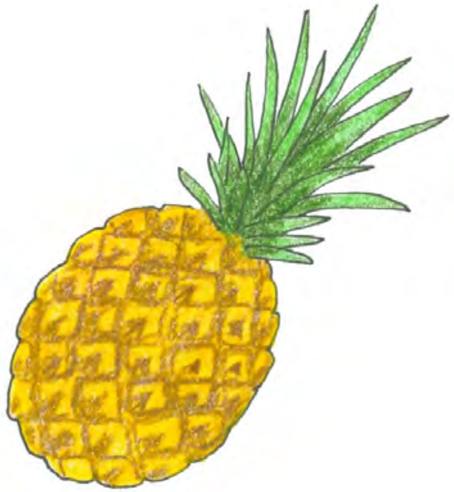


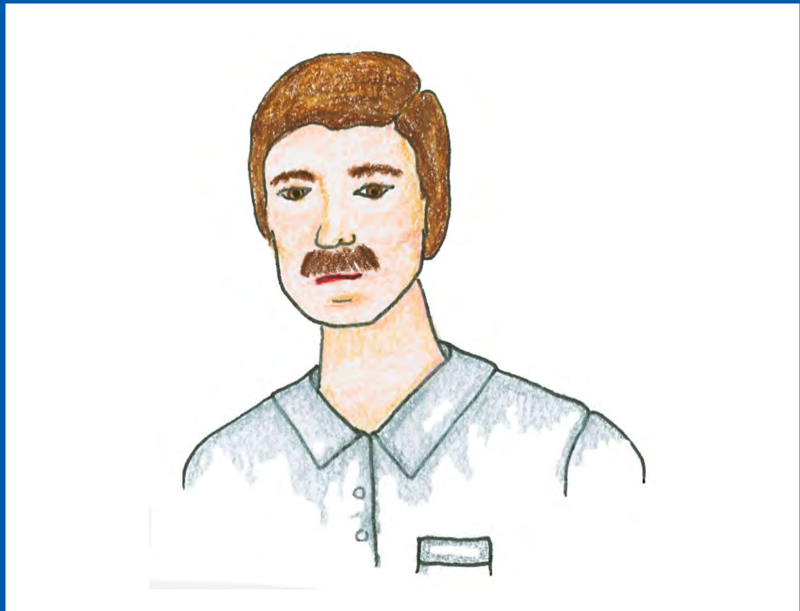
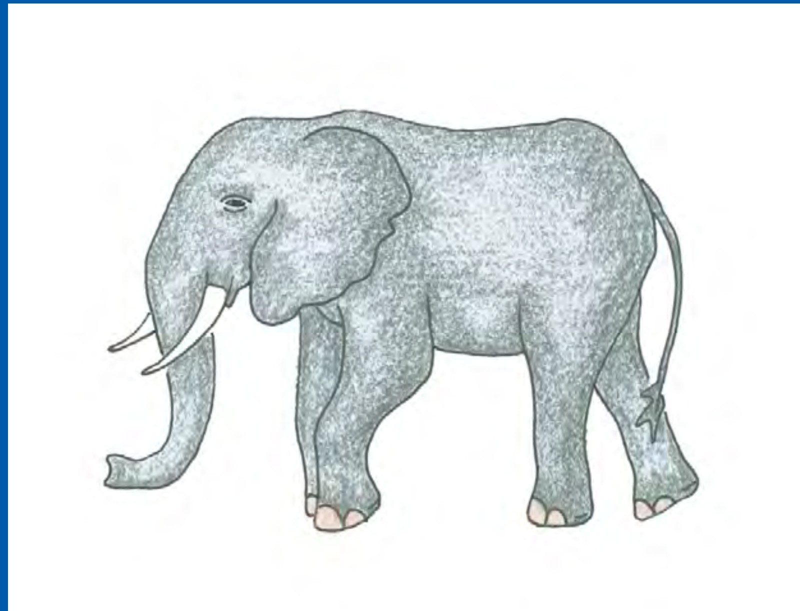
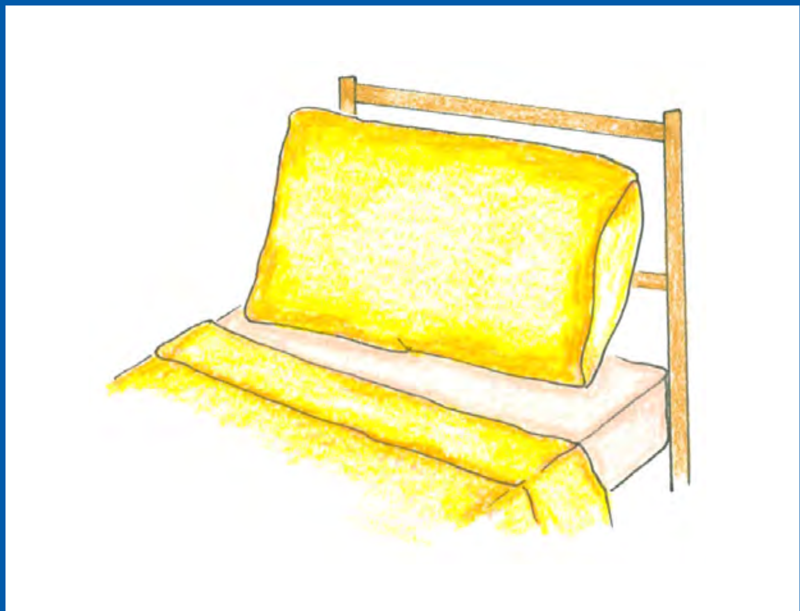
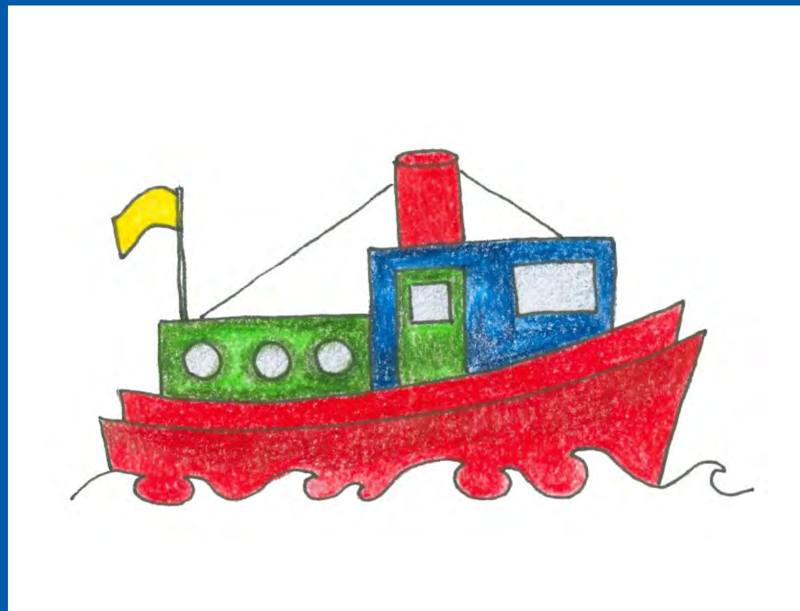






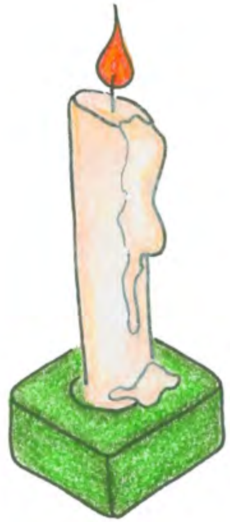
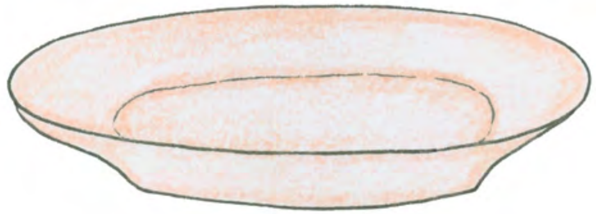



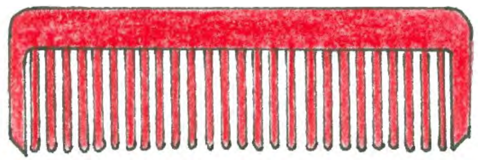


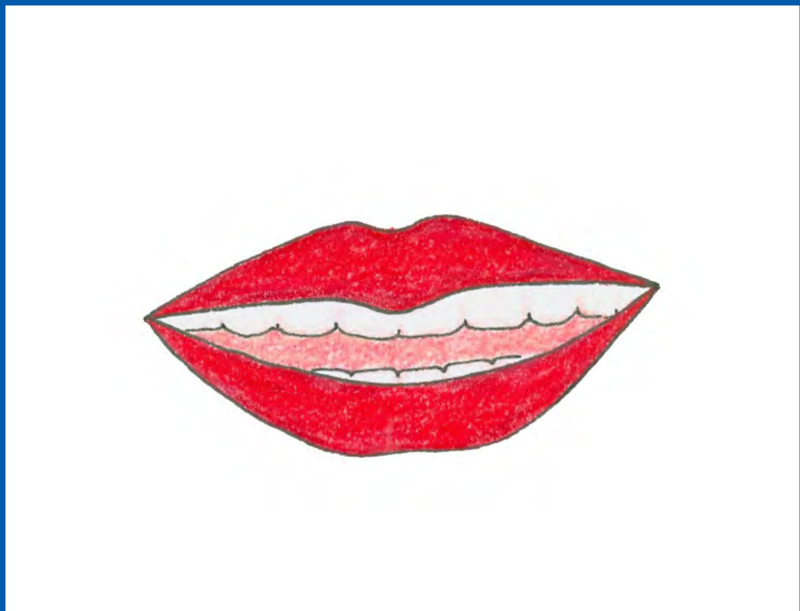


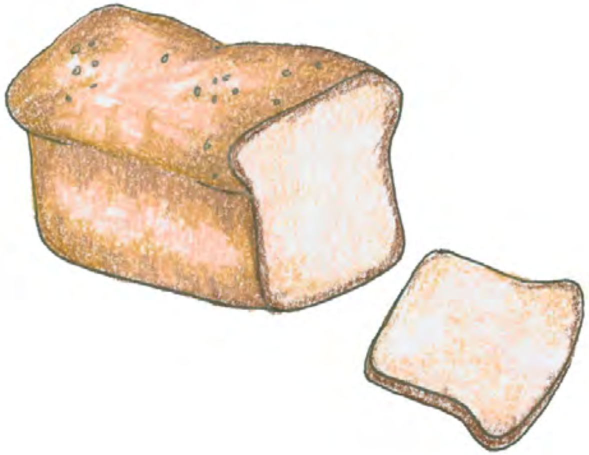


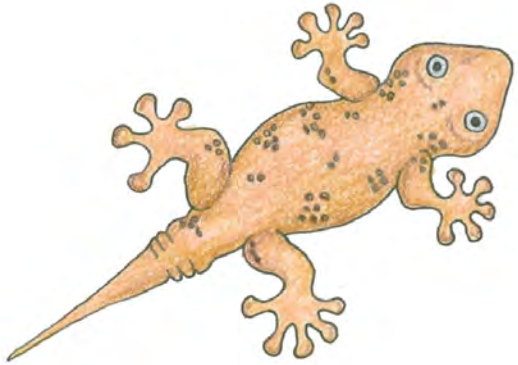




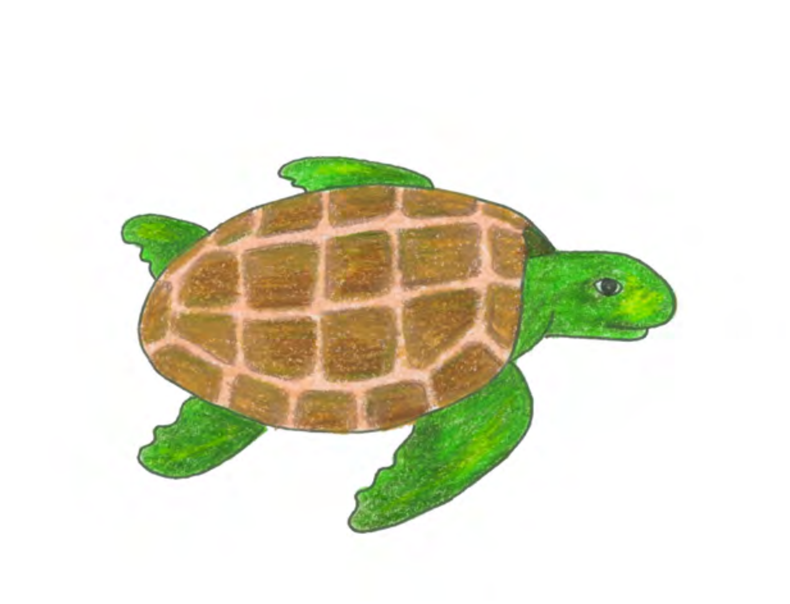
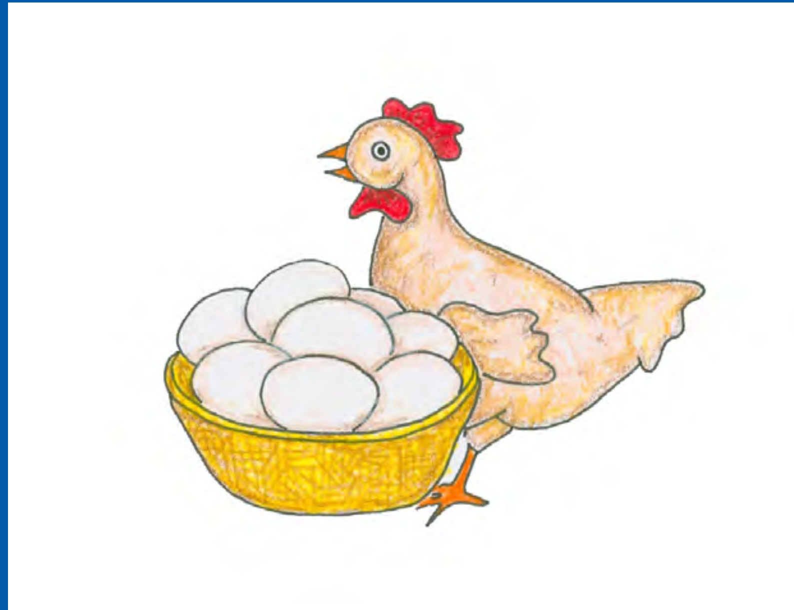


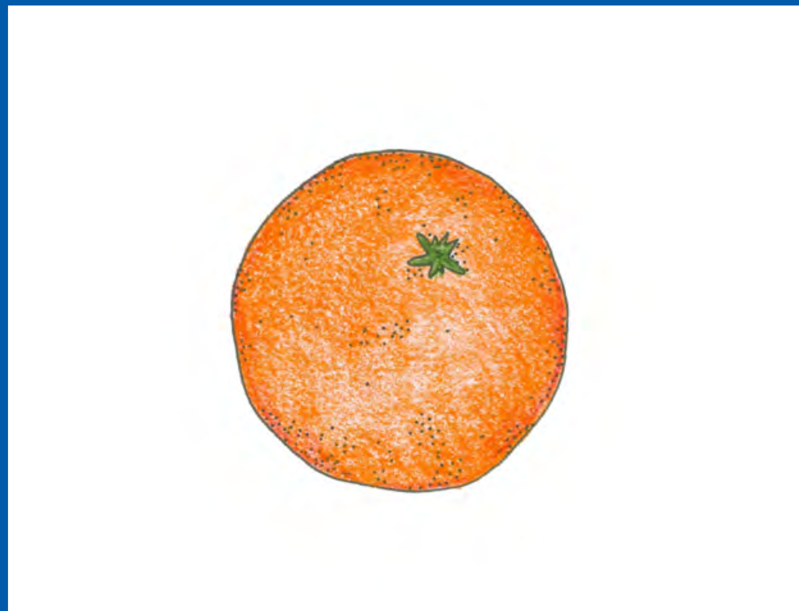




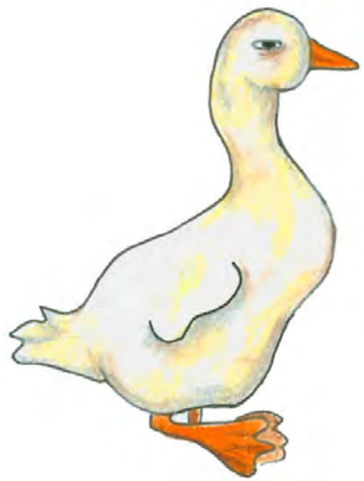


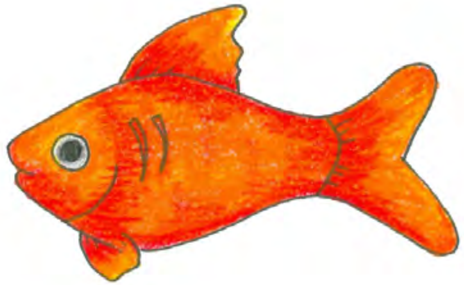



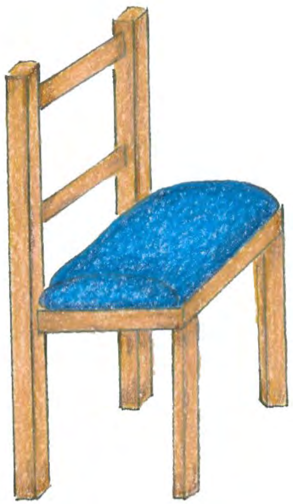


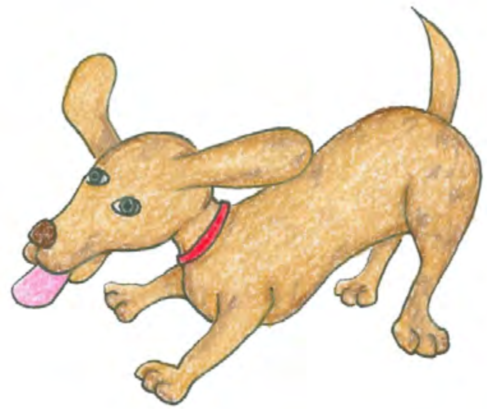


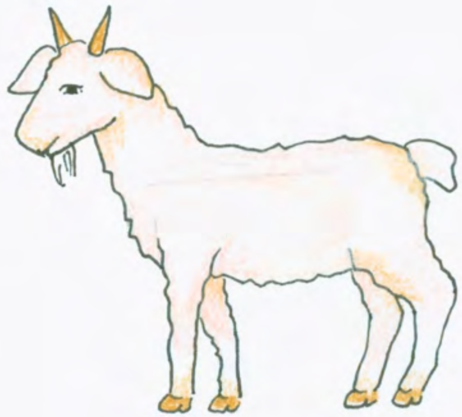



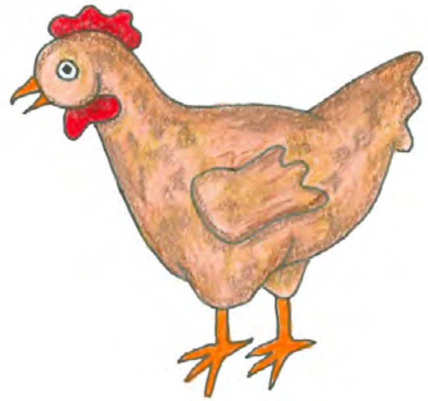


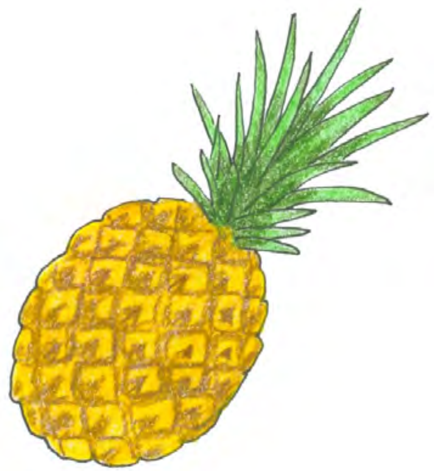


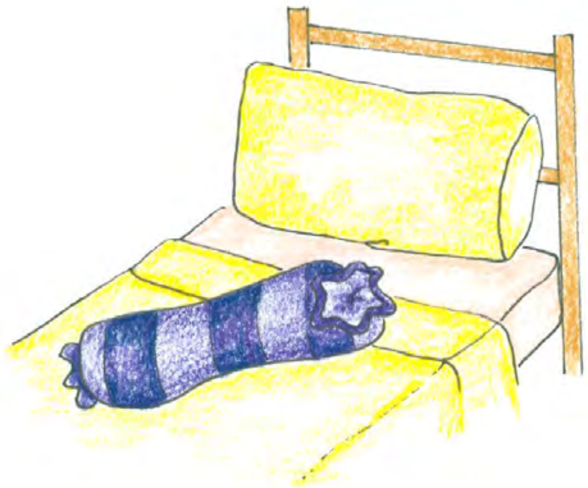
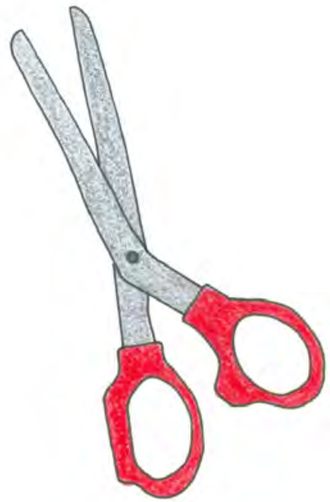



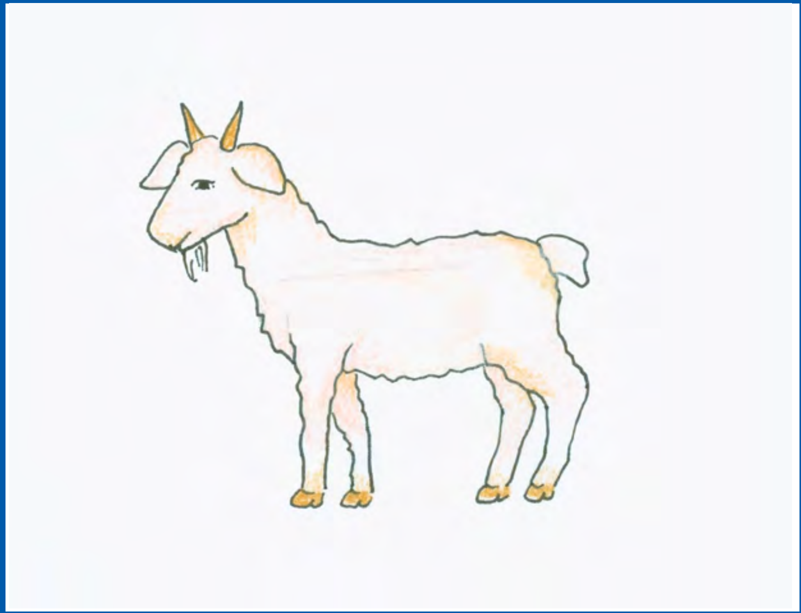



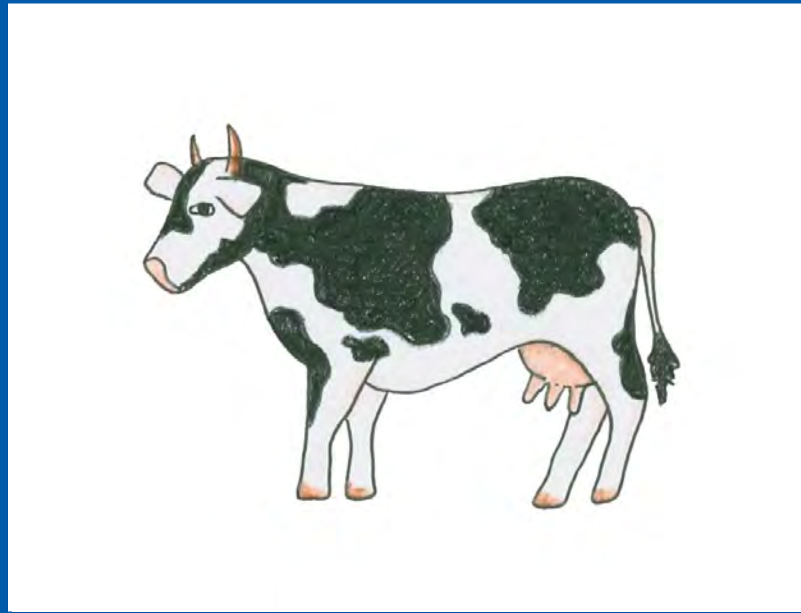


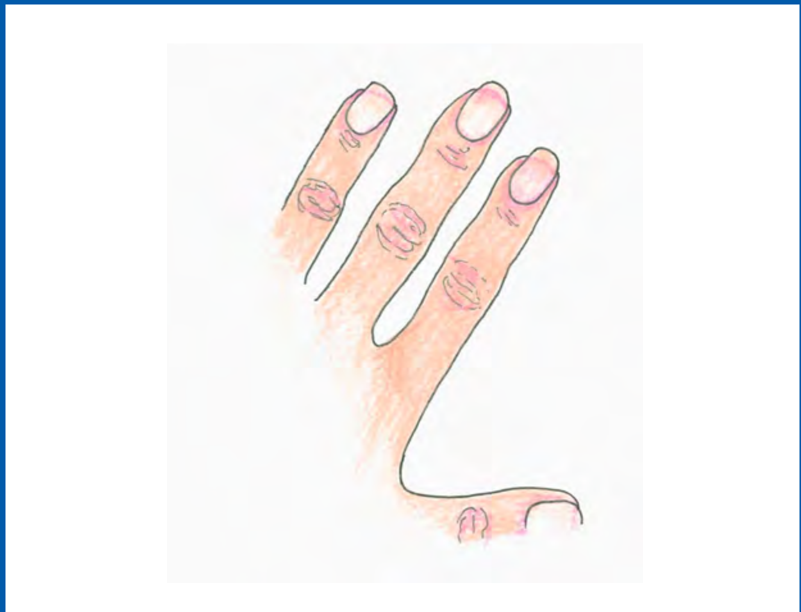


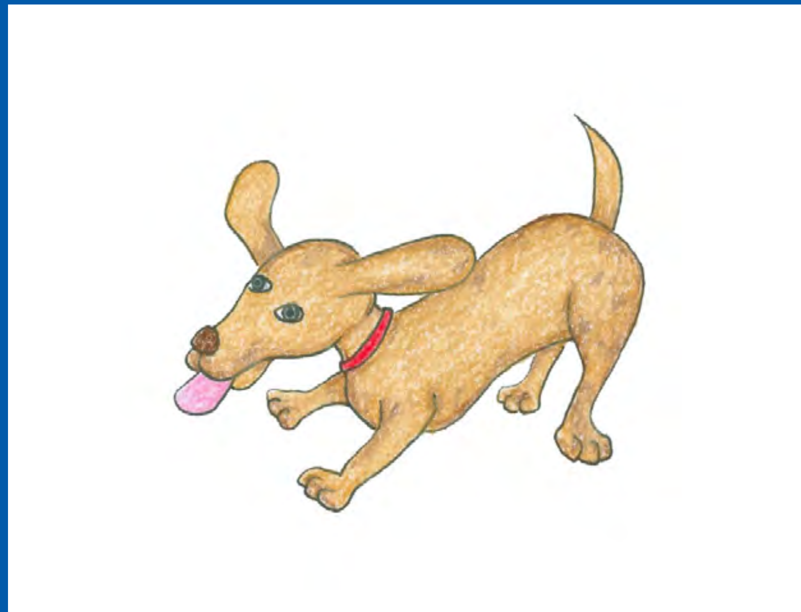


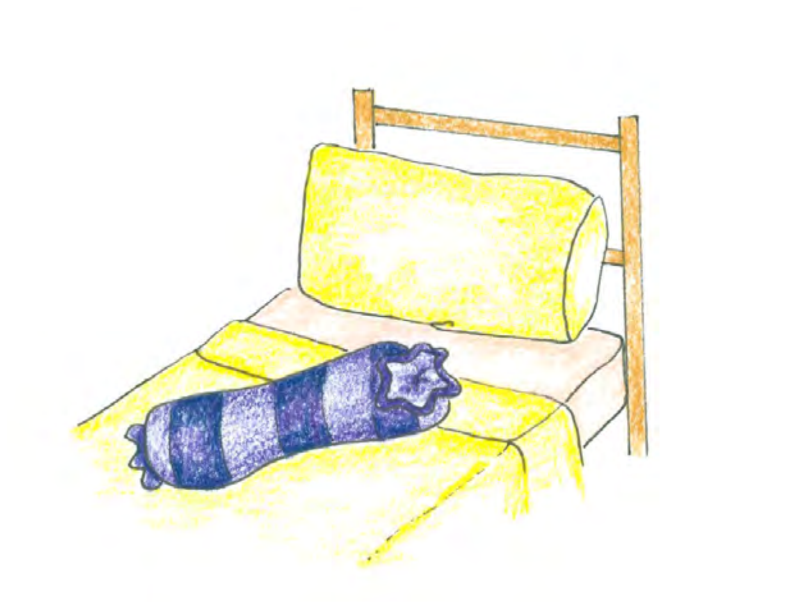


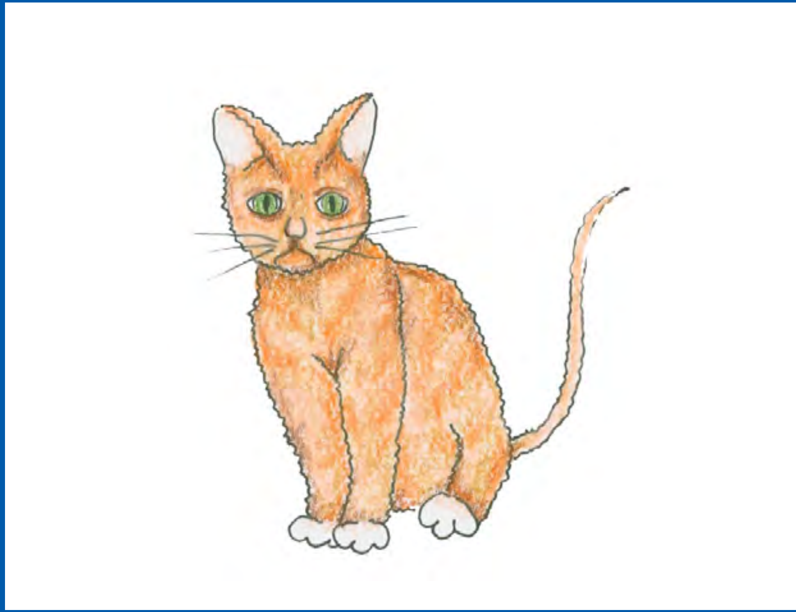


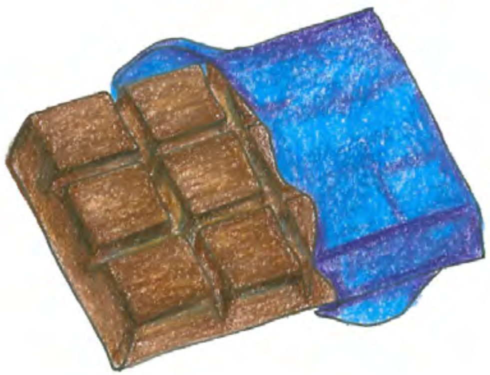
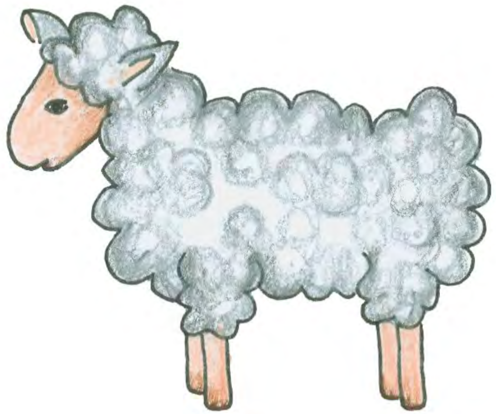
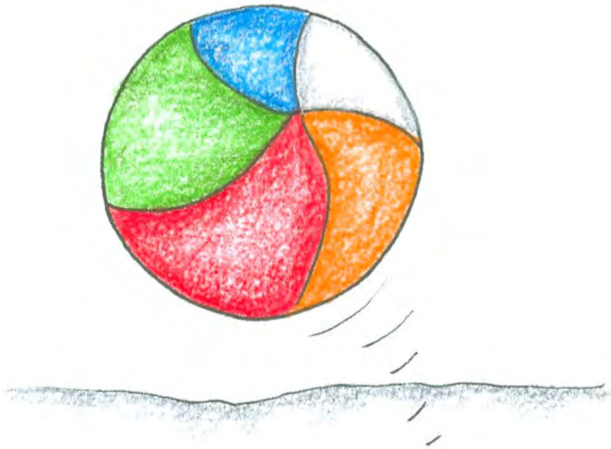


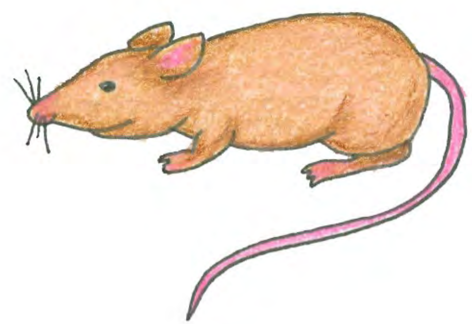


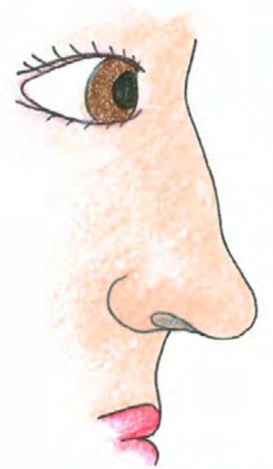


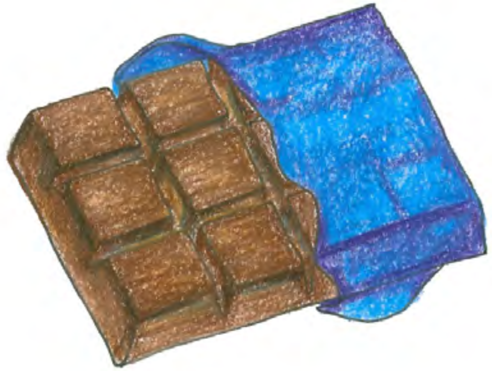



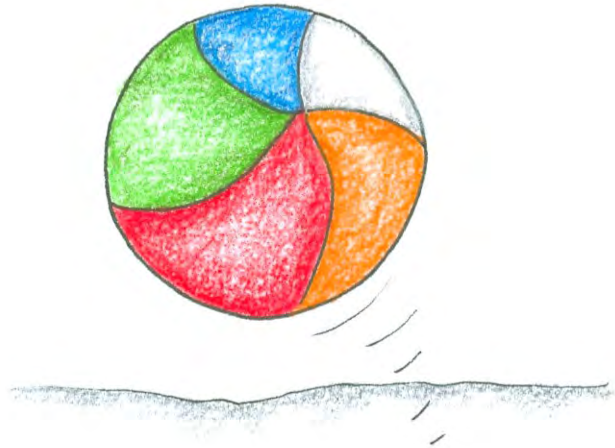


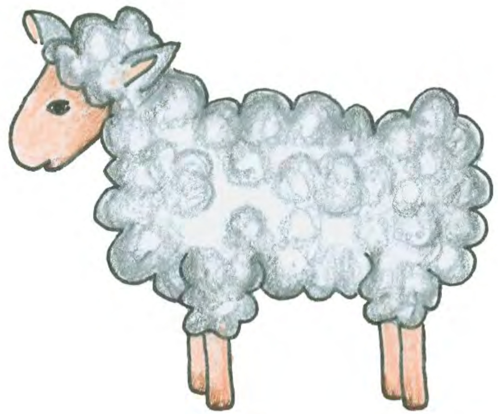




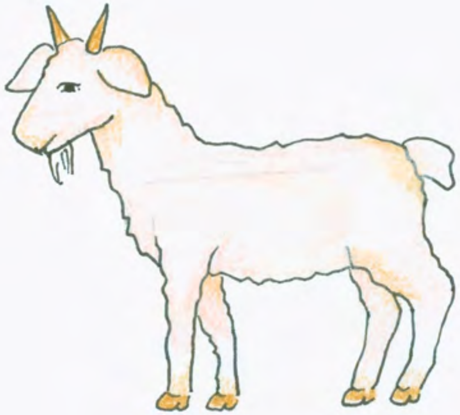


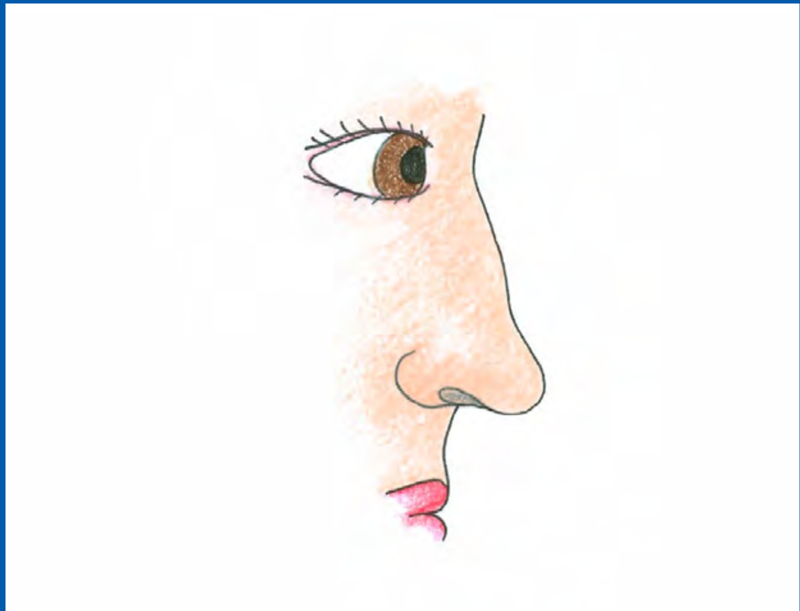


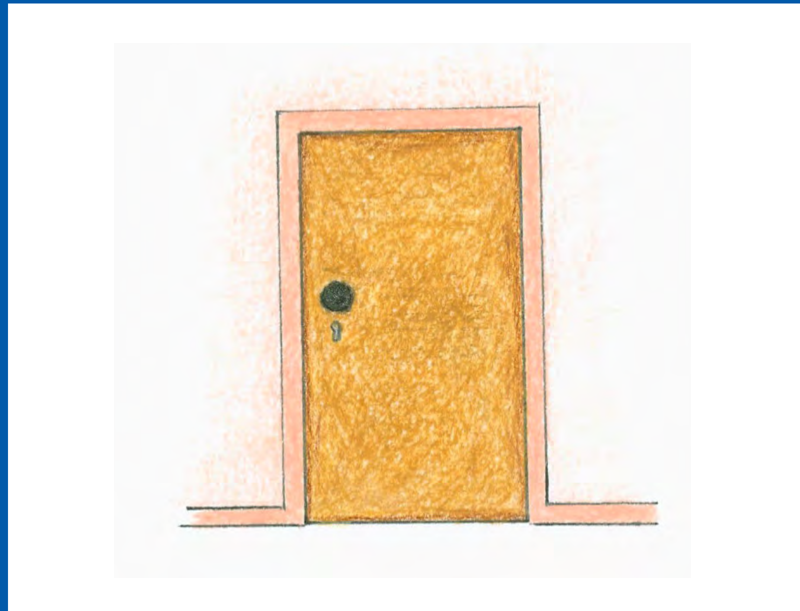


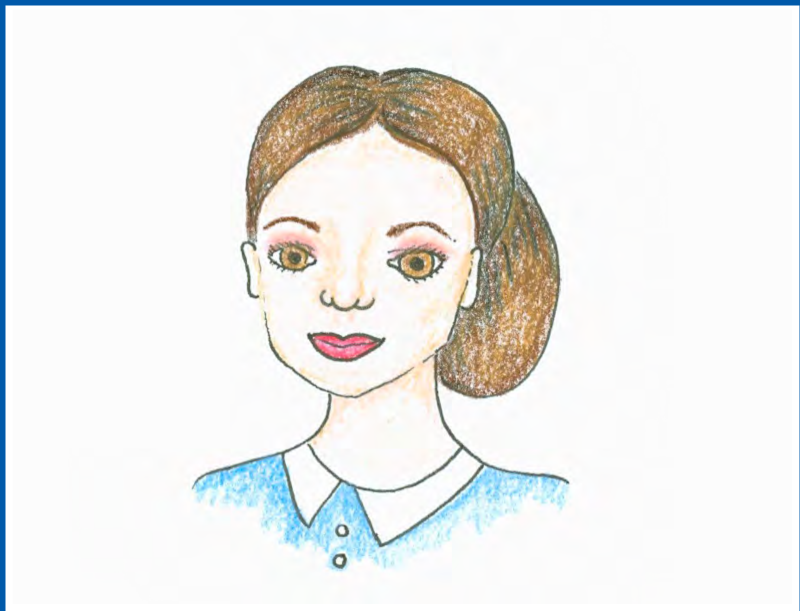


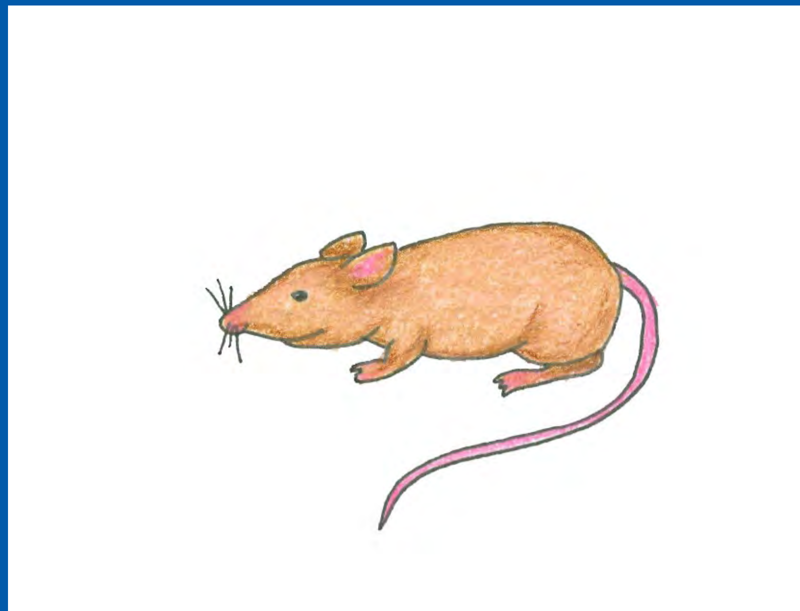


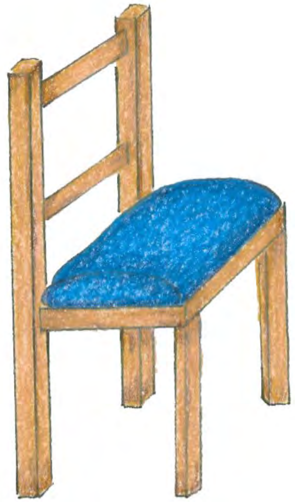


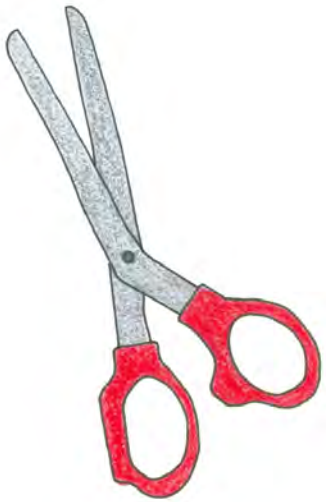



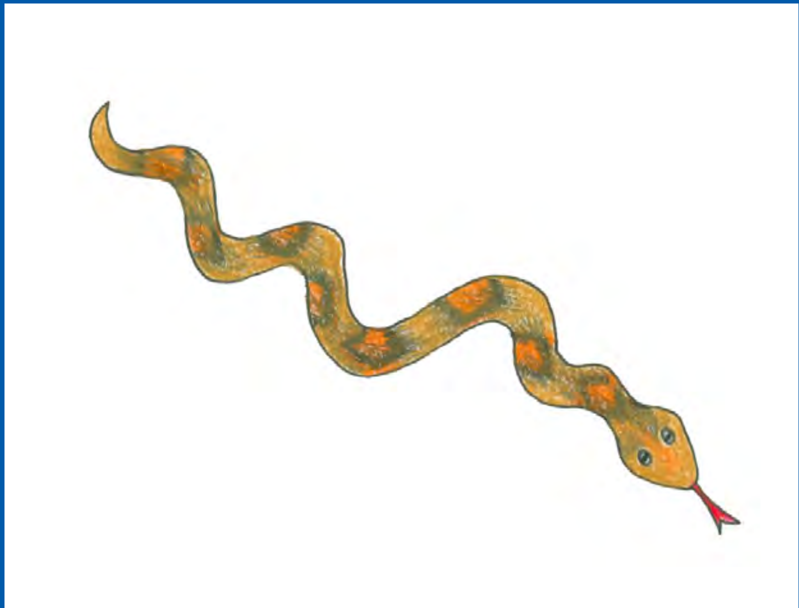


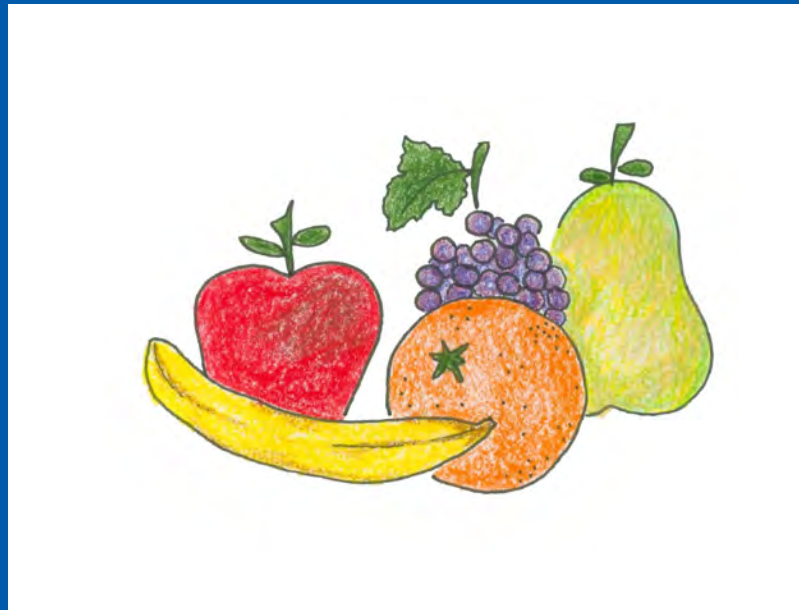


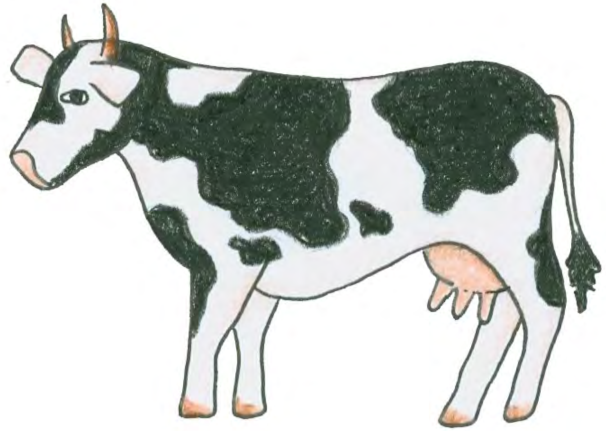
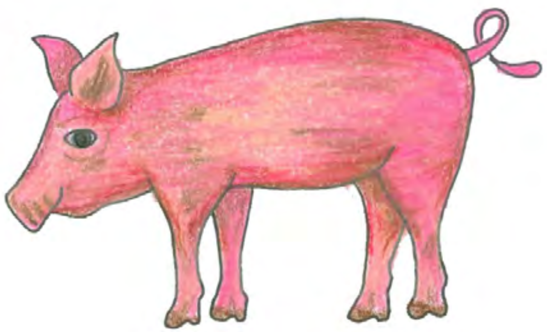









INDO-SPRITT

Indonesian – Speech Recognition Threshold Test
(Tes ambang pendengaran dengan tutur kata bahasa Indonesia)

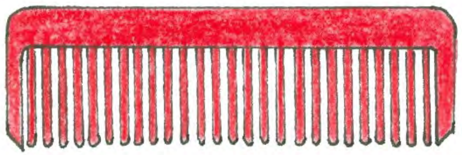
Book B (Test form 3 & 4)
Buku B (Susunan tes 3 & 4)

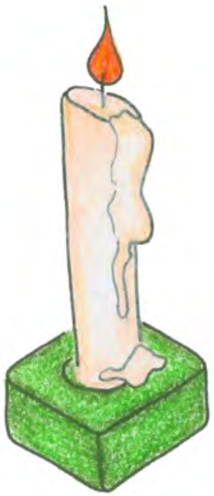
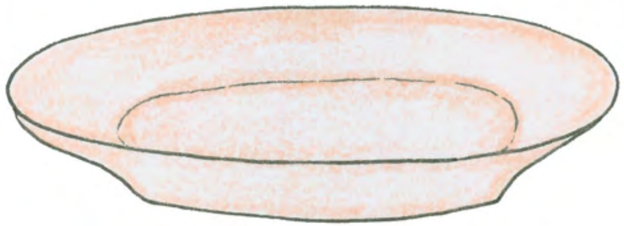
Dahlia E Sartika

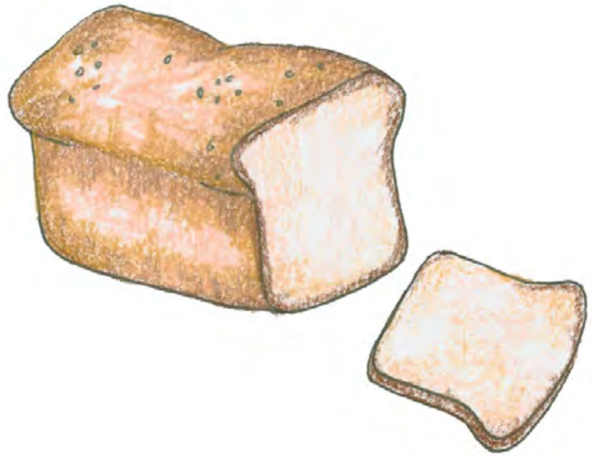
Philip Newall, Mridula Sharma, Robert Mannell, Harvey Dillon

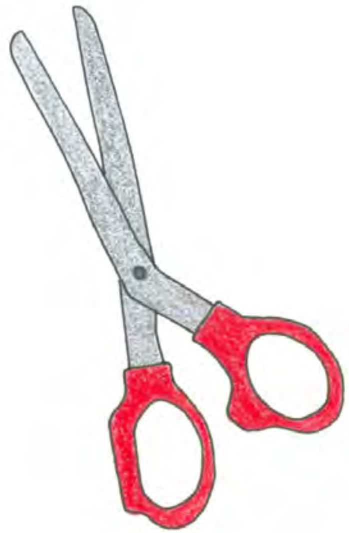
MACQUARIE UNIVERSITY
SYDNEY ~ AUSTRALIA

RIDBC
Renwick Centre
For Research and Professional Education



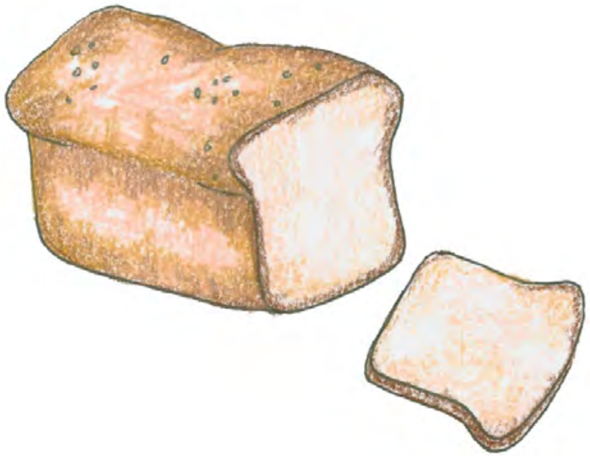


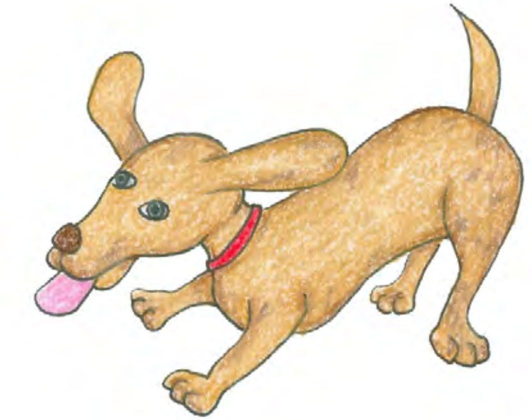


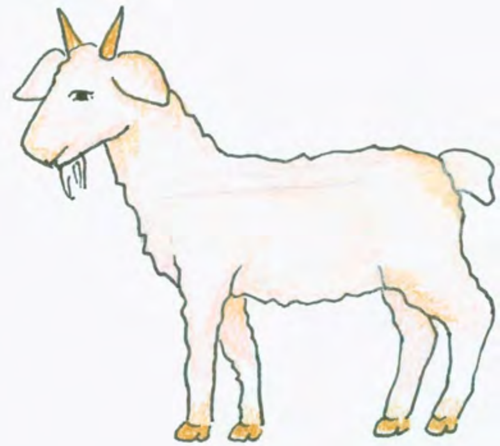


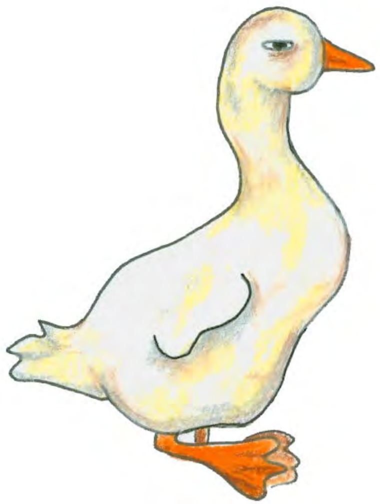


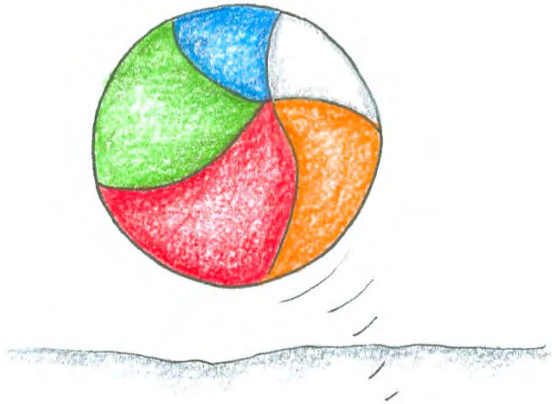


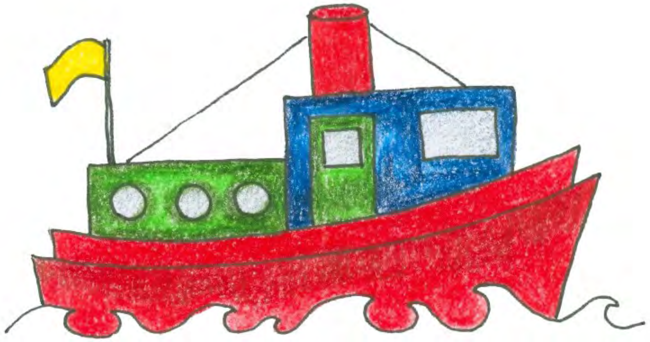
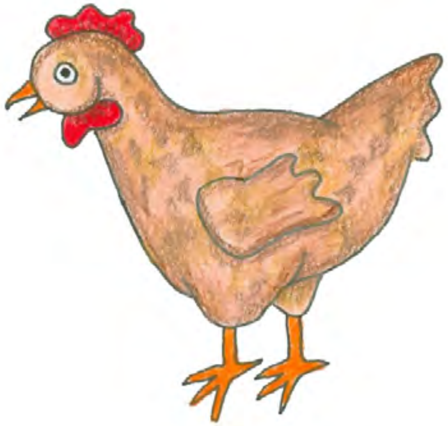


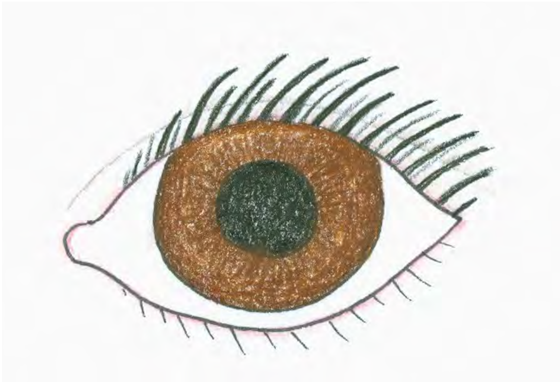
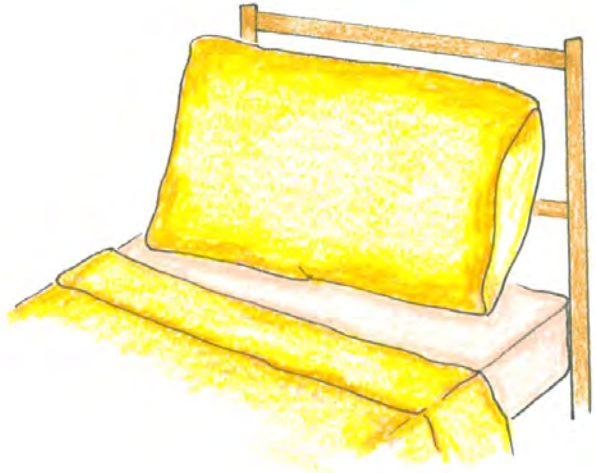
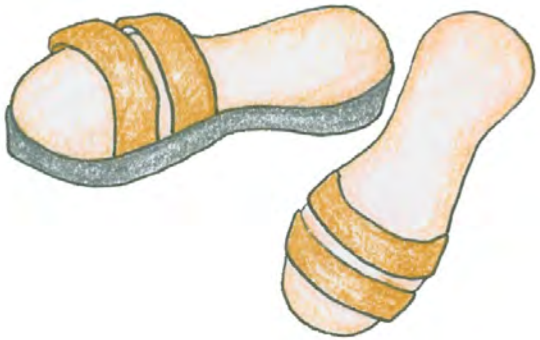
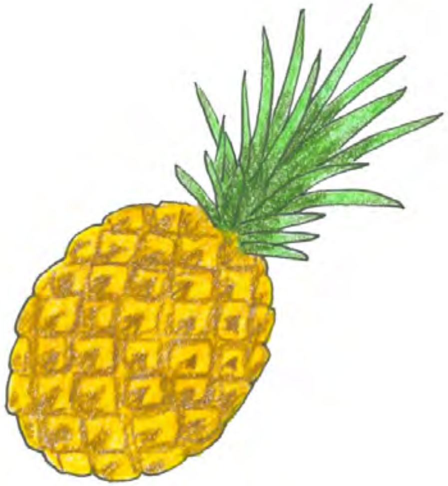


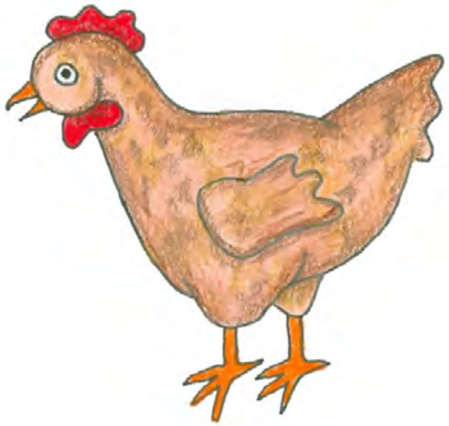


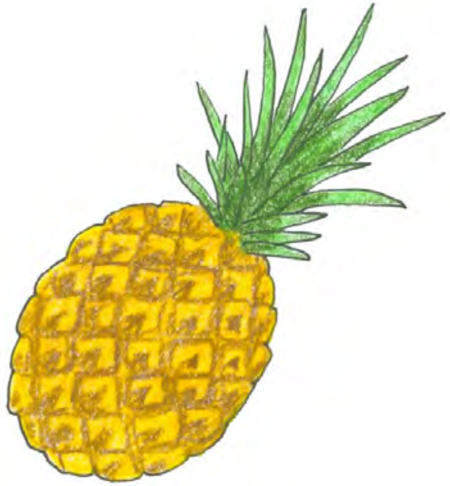


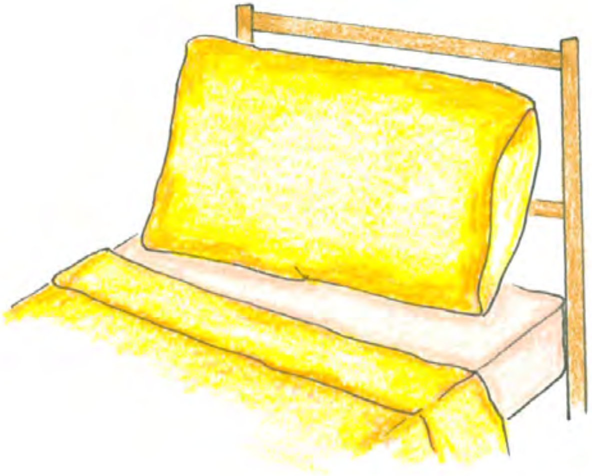
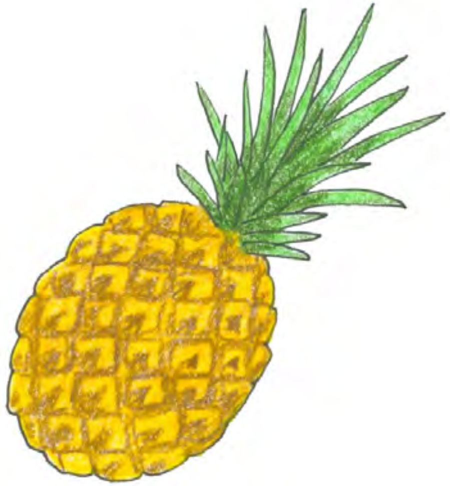
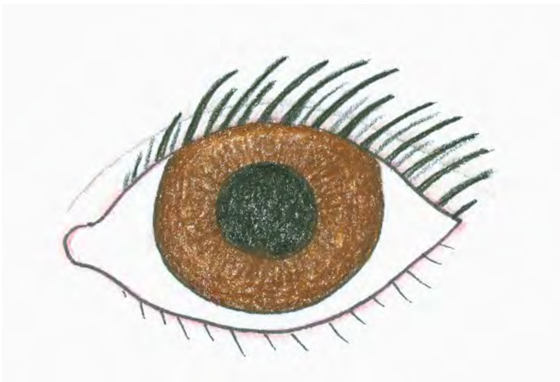



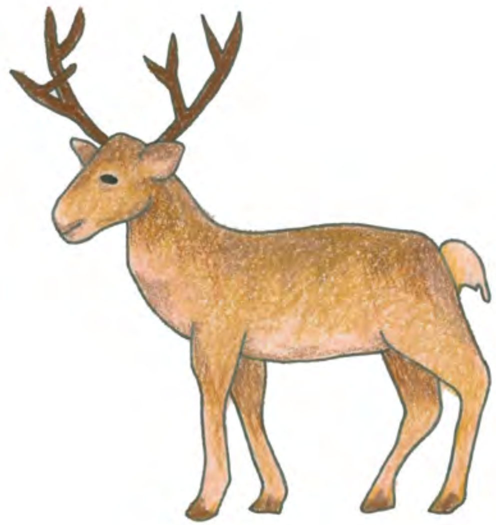





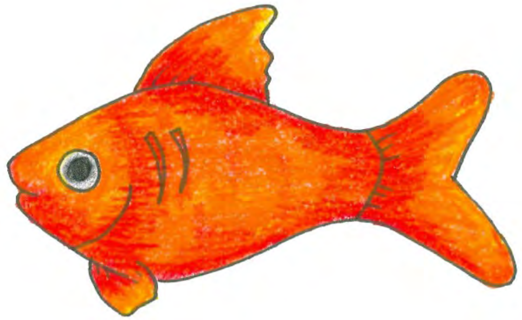


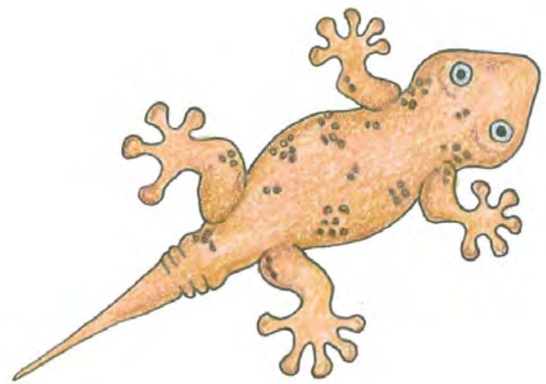



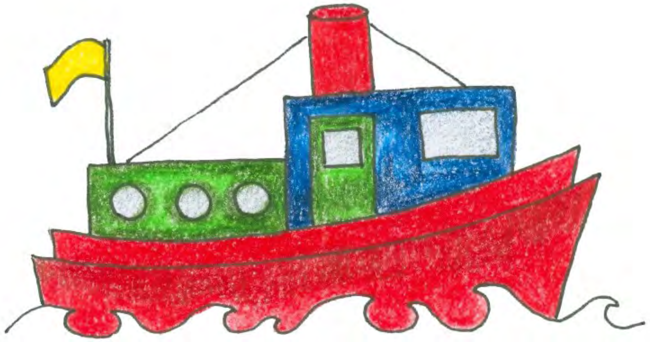


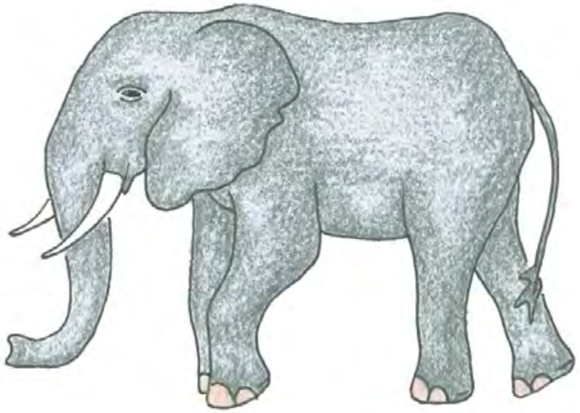





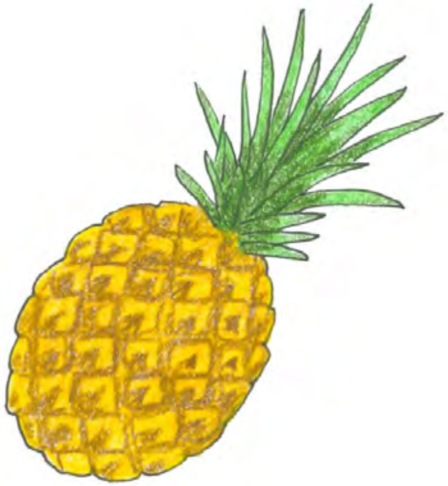



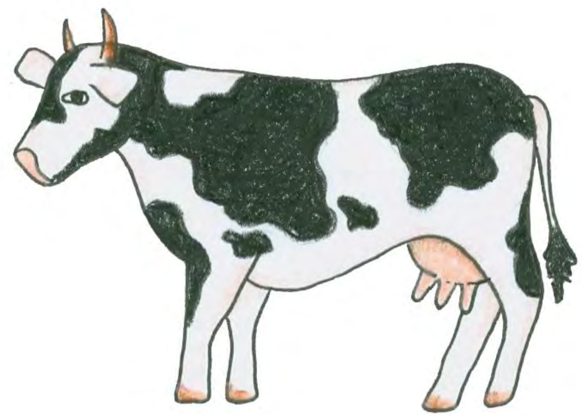




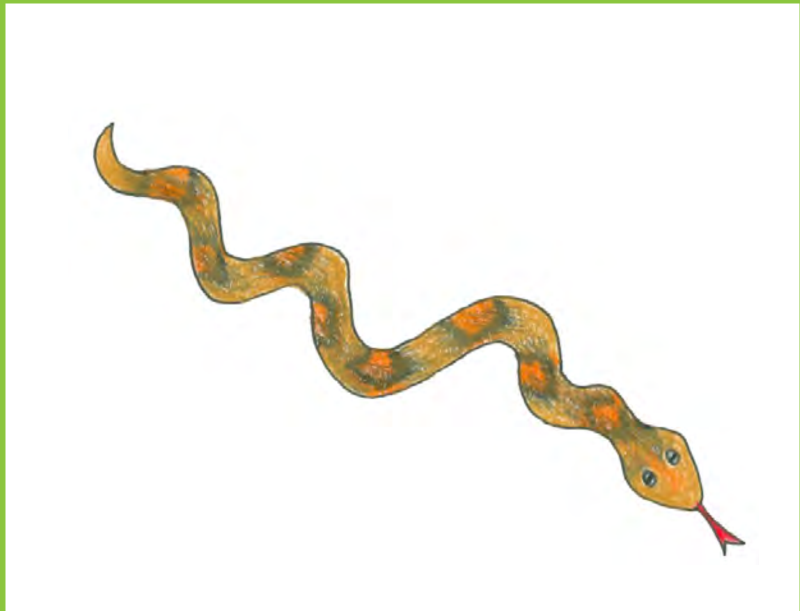


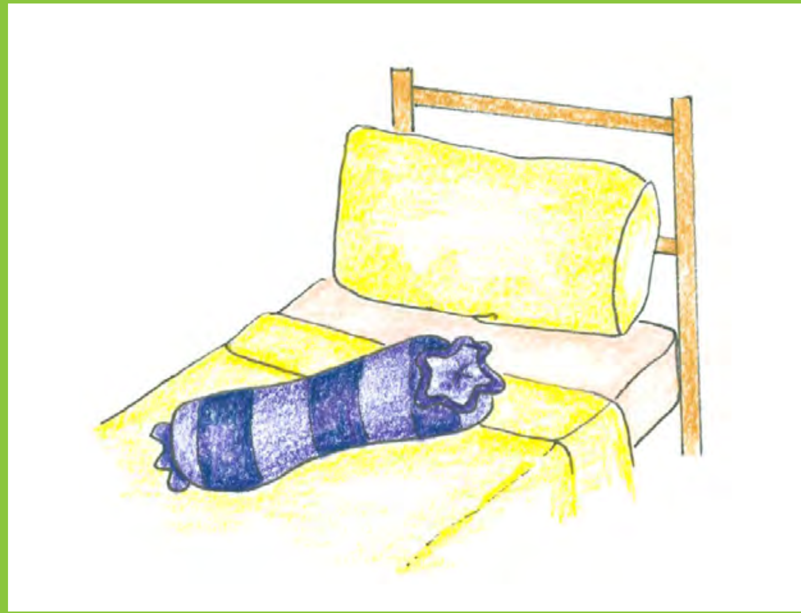


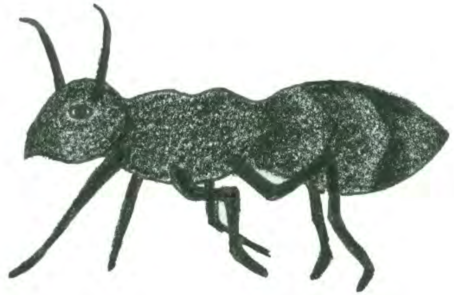


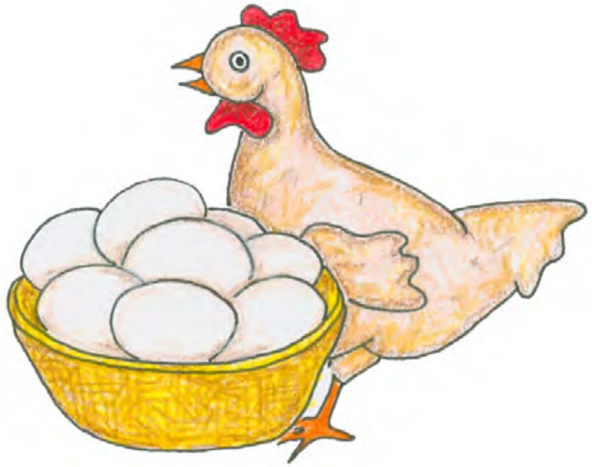


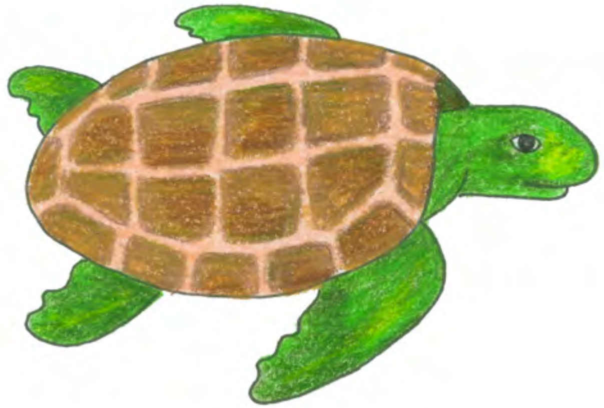


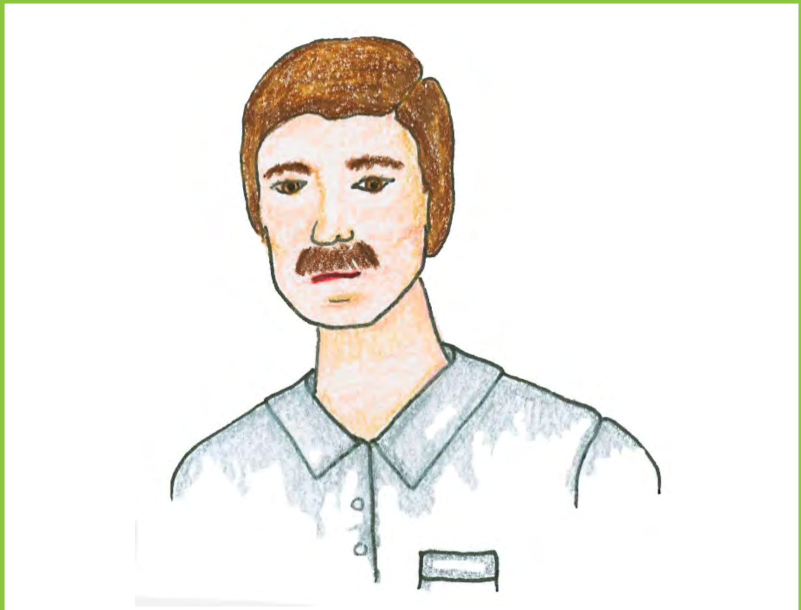



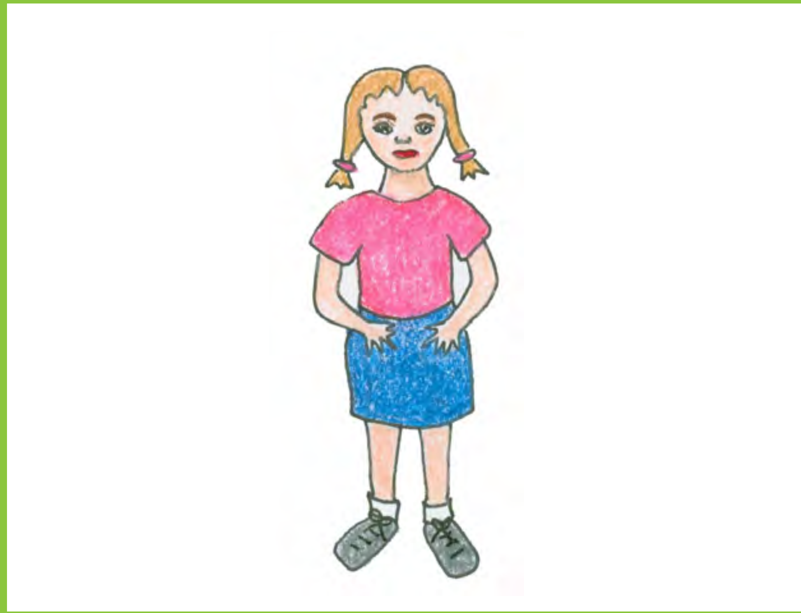



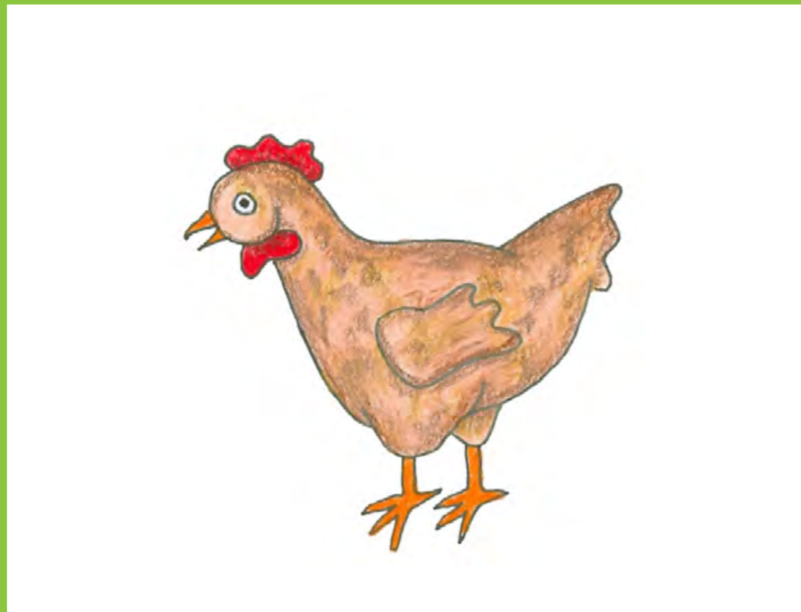


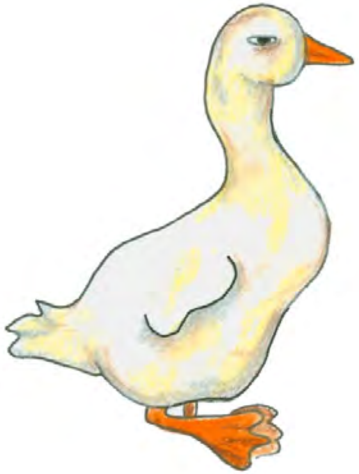


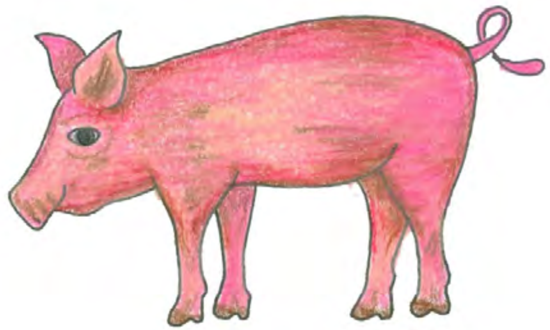
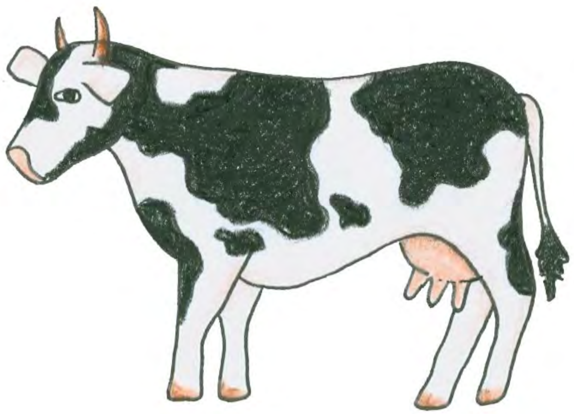


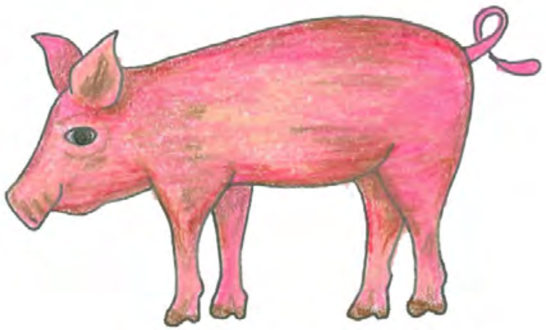


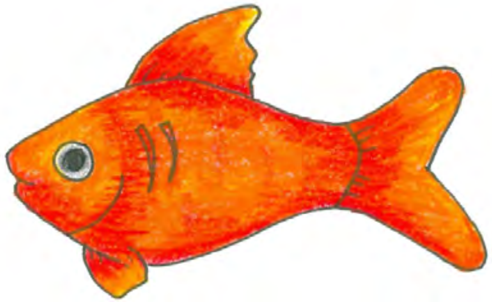




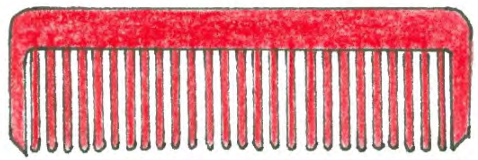
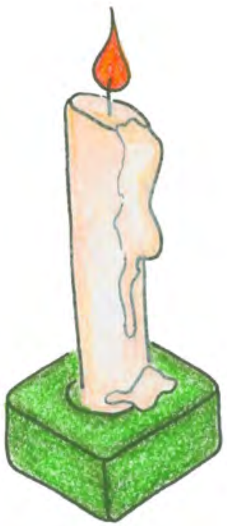
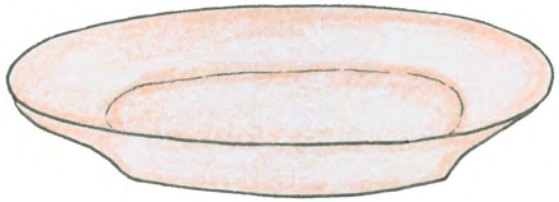



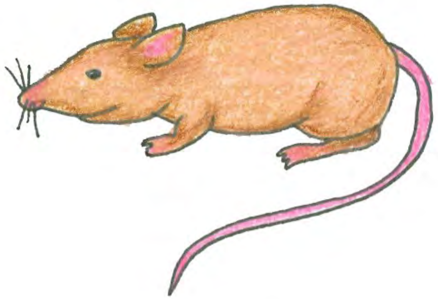



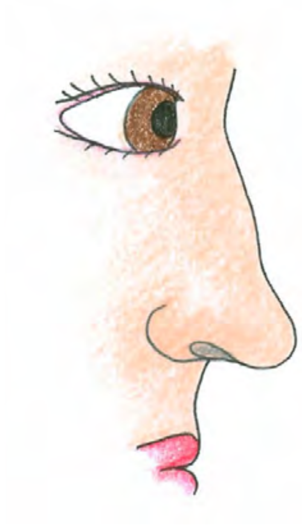


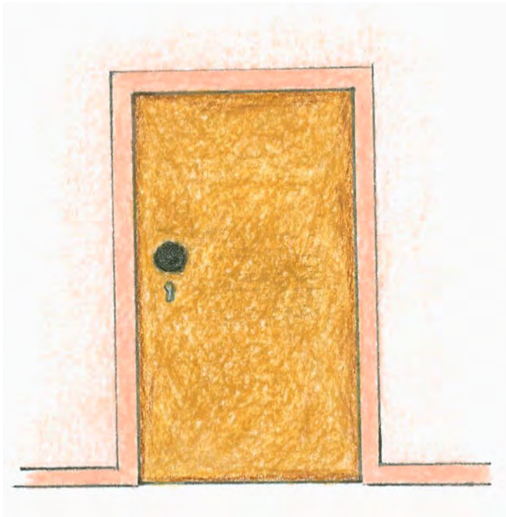


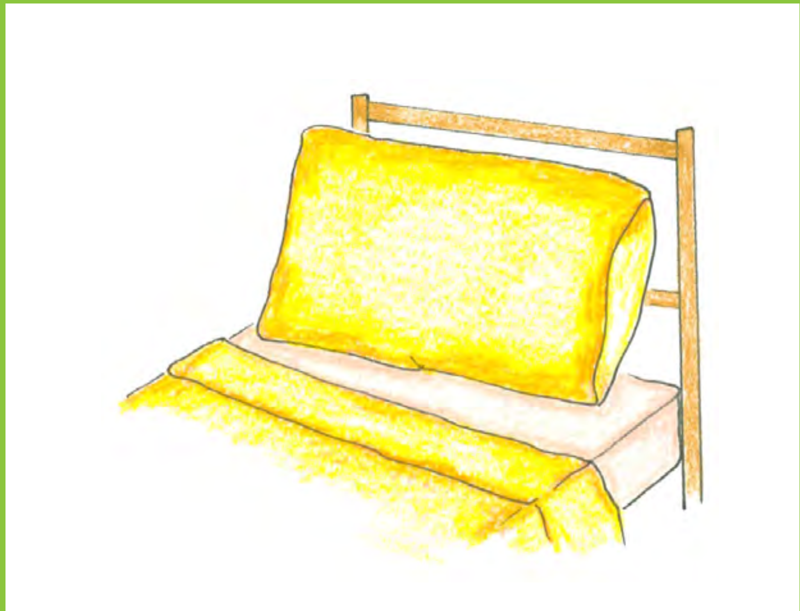
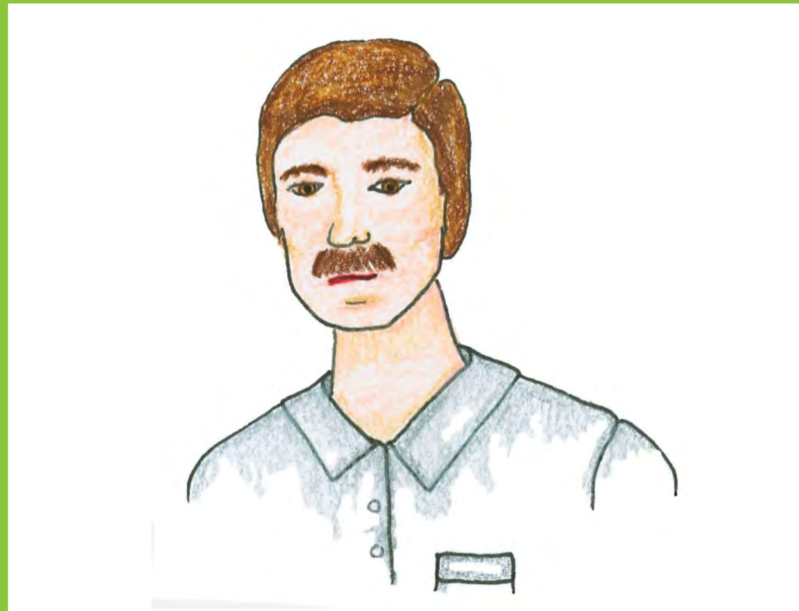
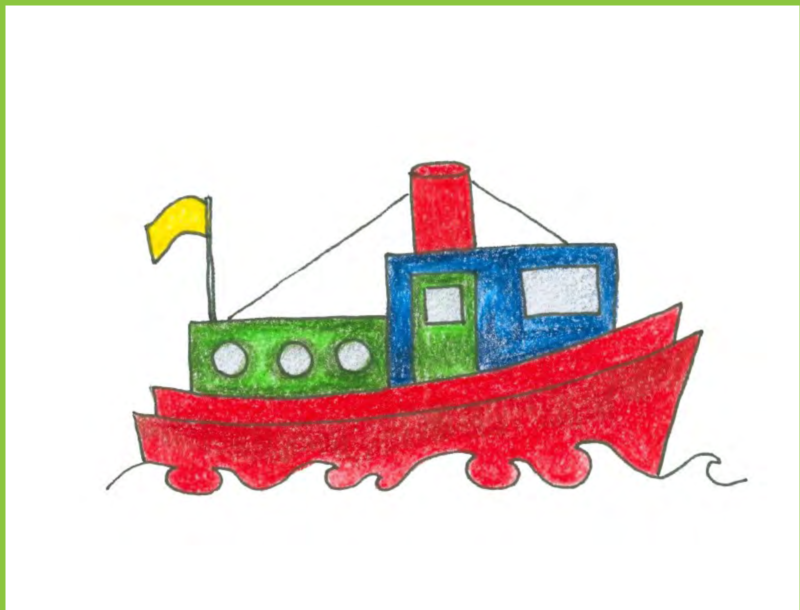
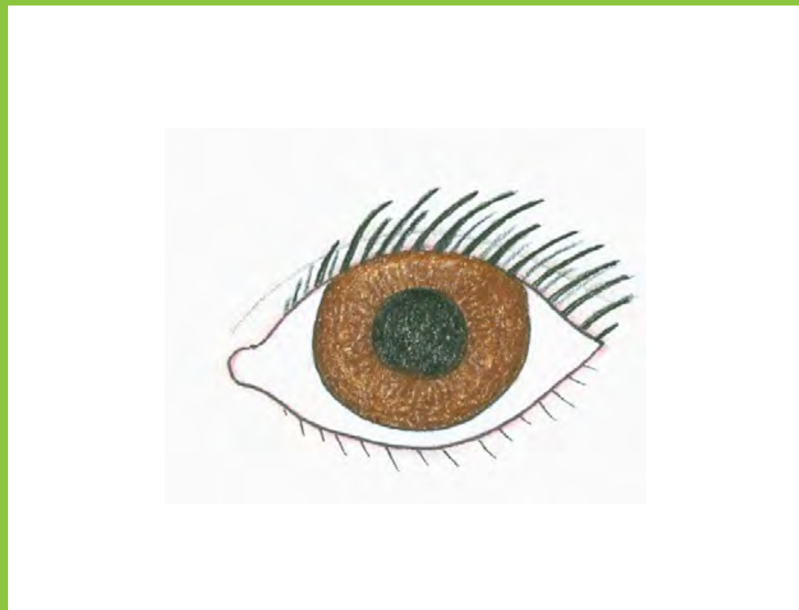




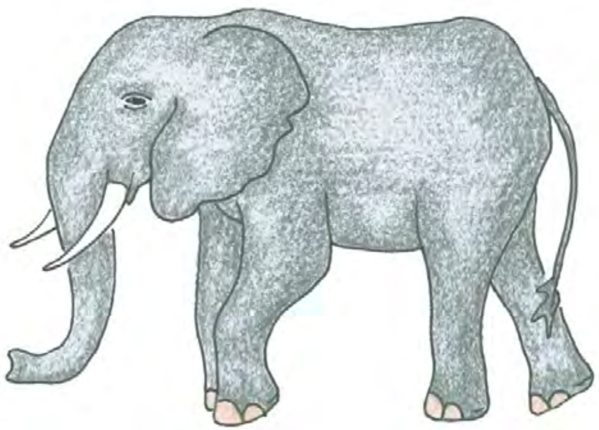


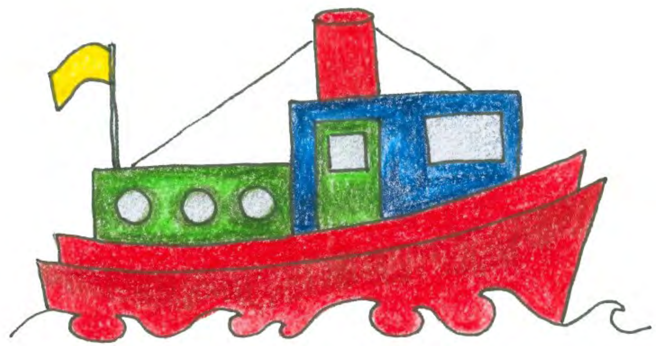



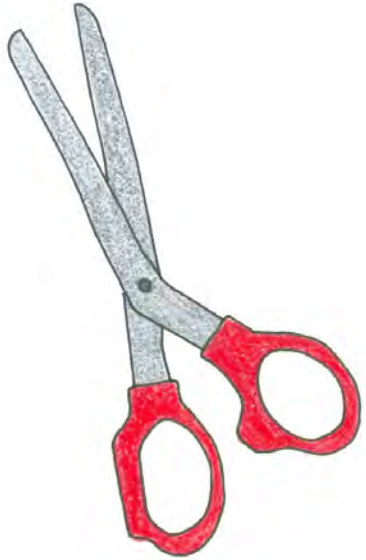


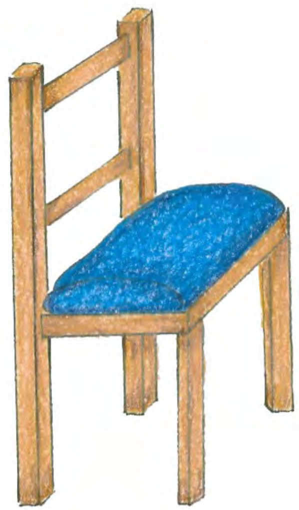


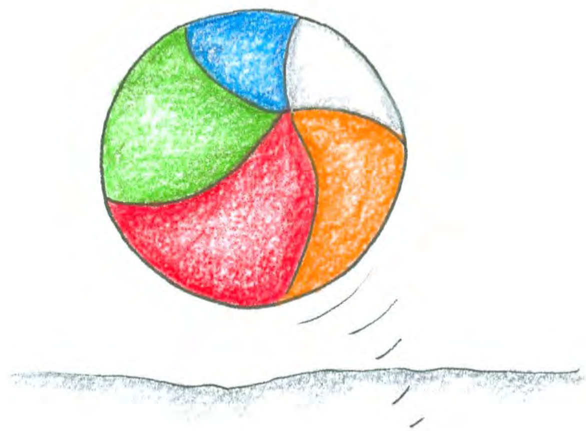







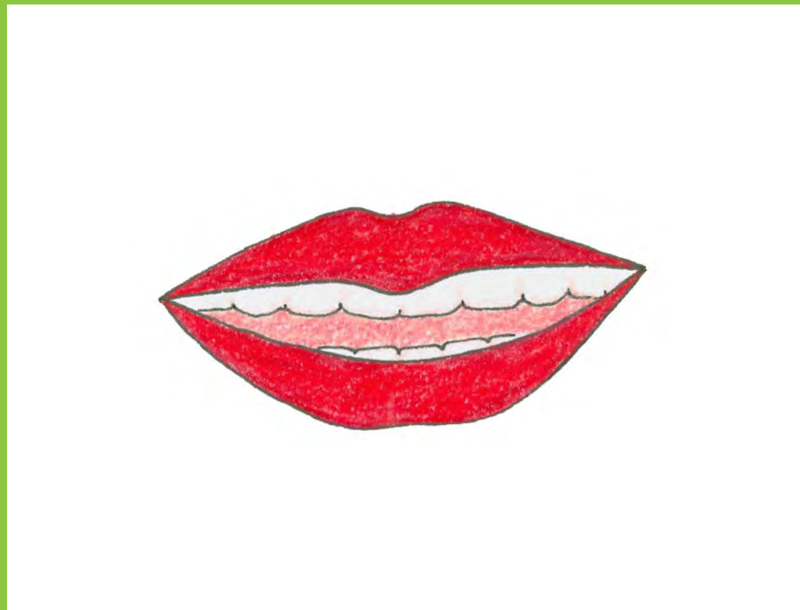


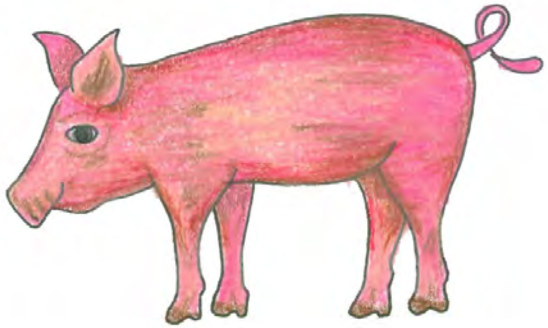


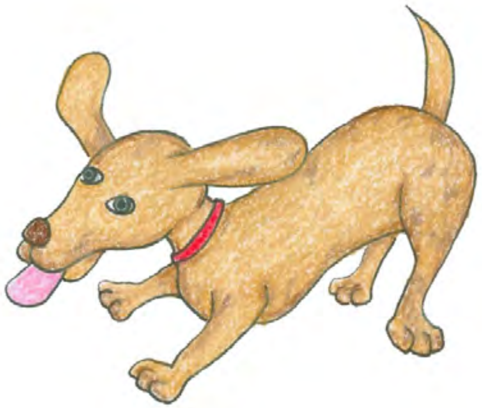


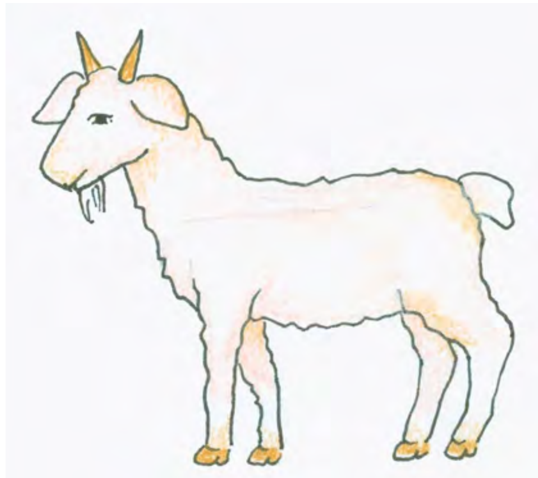


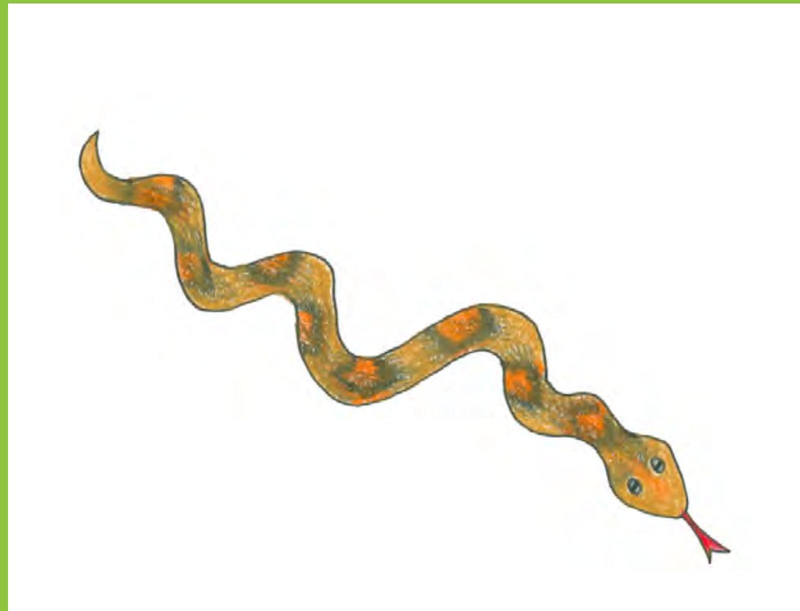


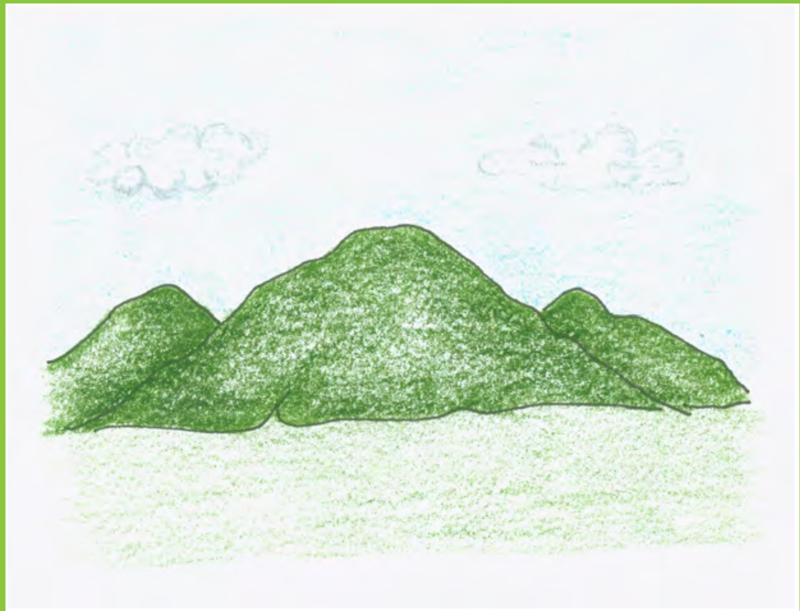



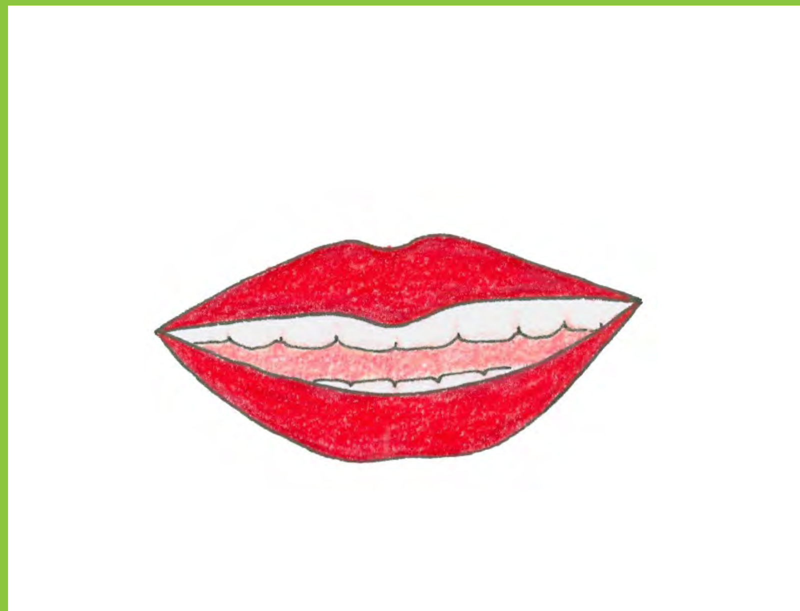


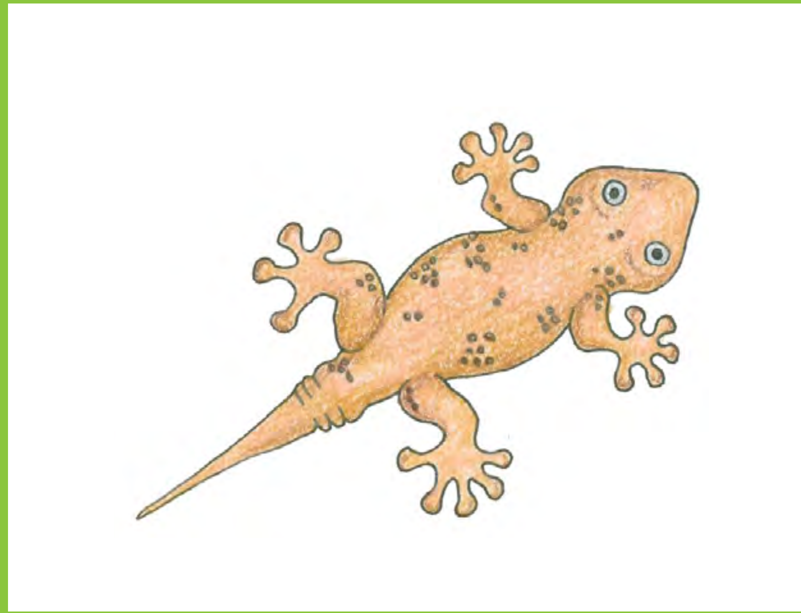




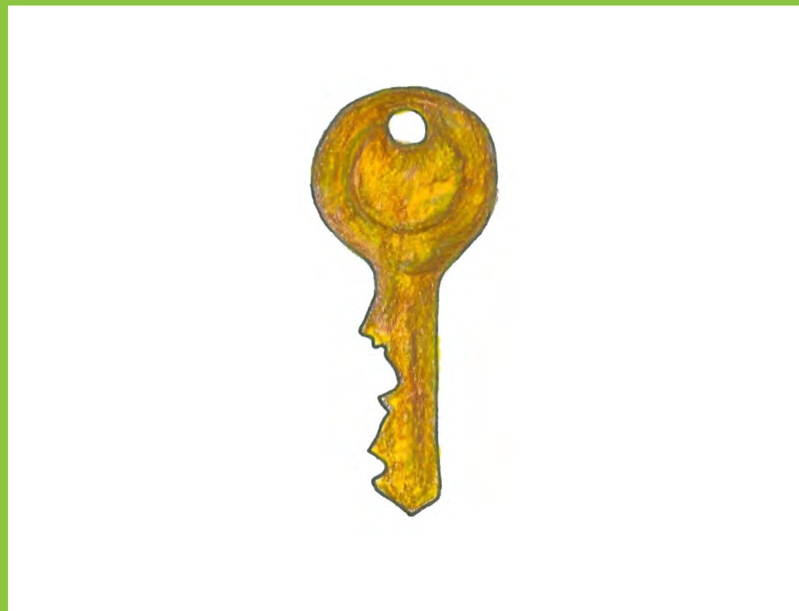


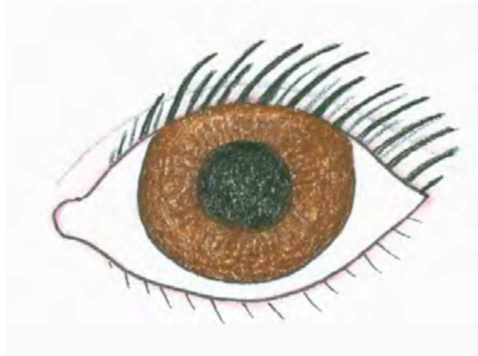
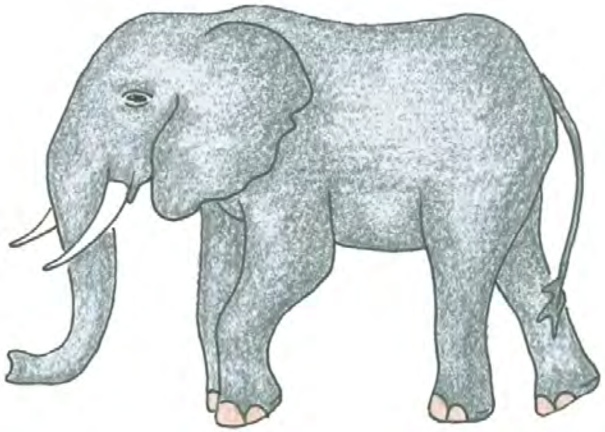
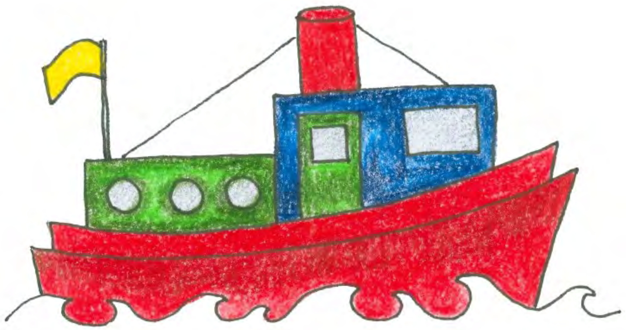


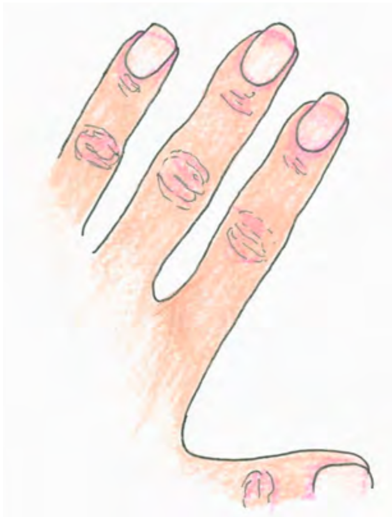


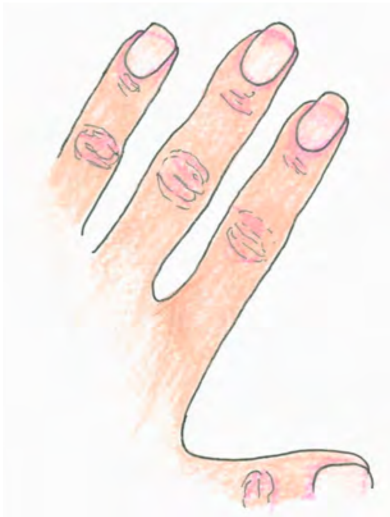


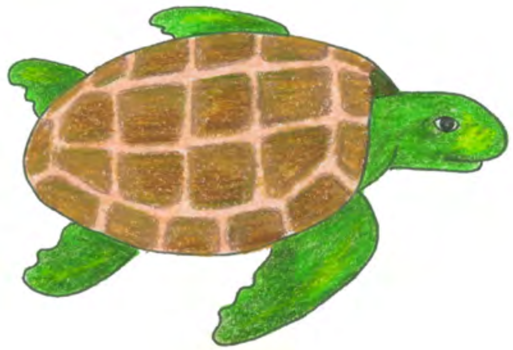
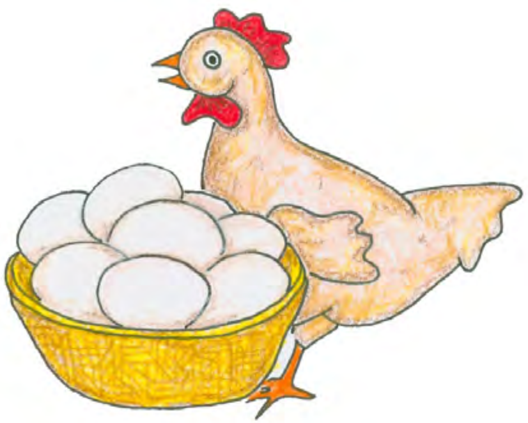



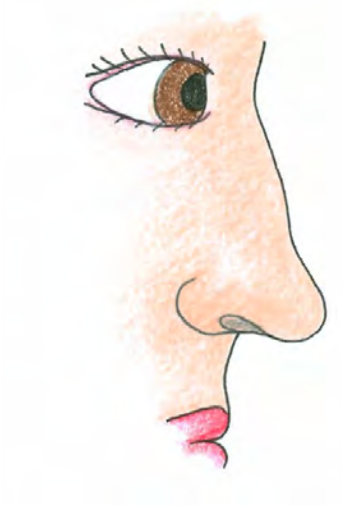
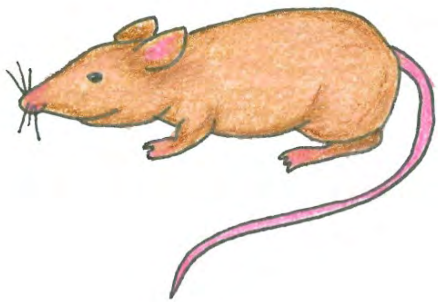
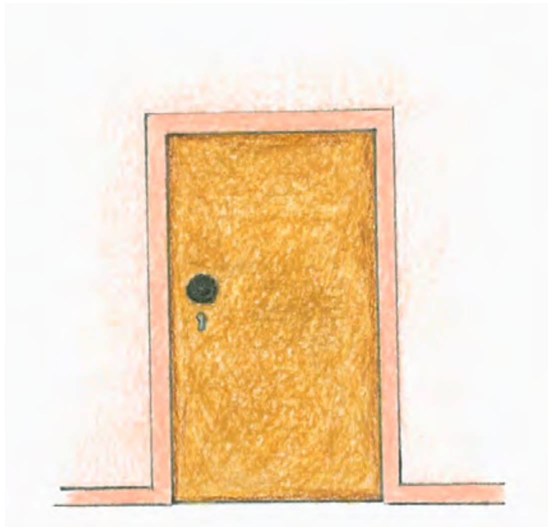


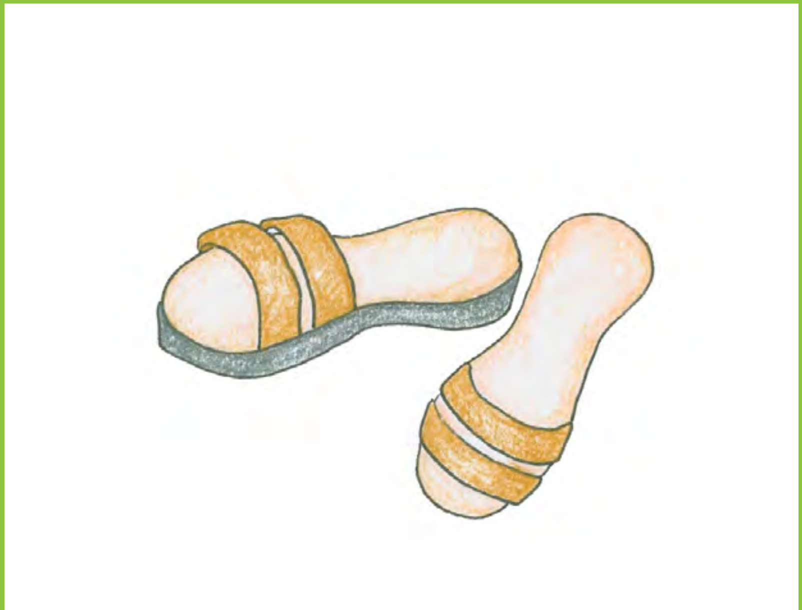



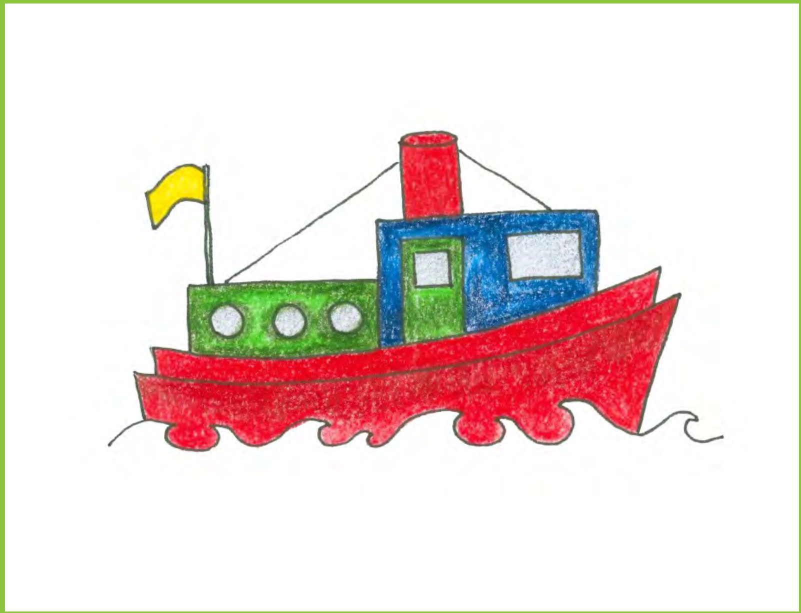


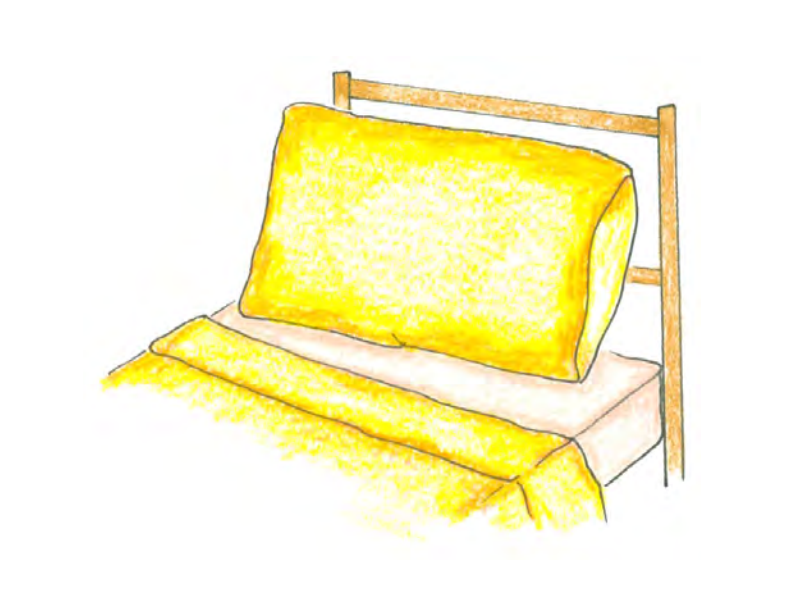




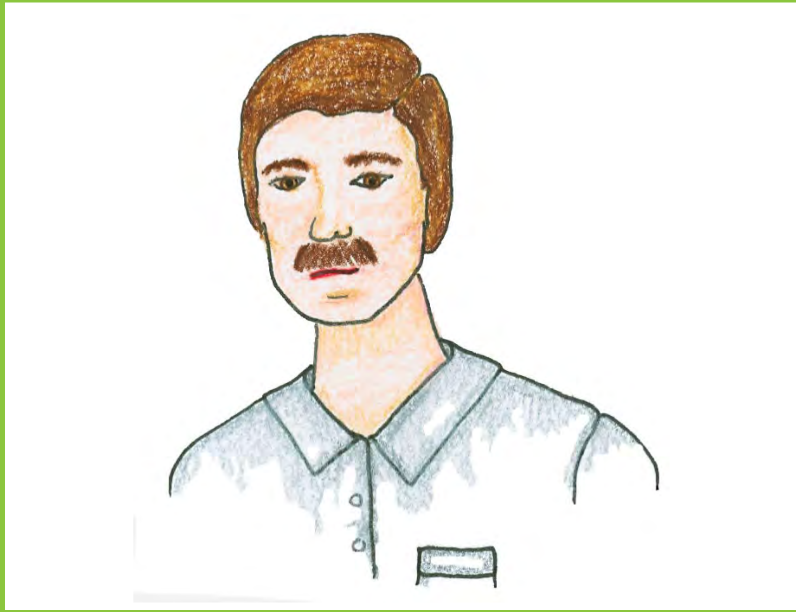


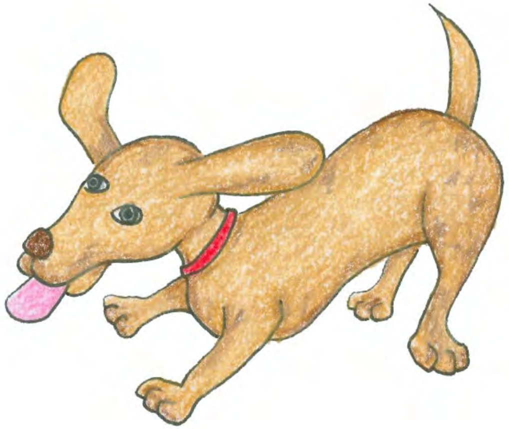
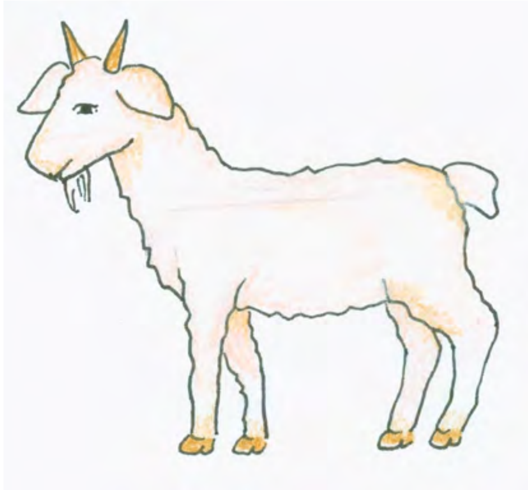












APPENDIX F: ADAPTIVE TEST FORMS

Test Form A1

Item no.	Words	Excursions
1	Bantal	
2	Kaki	
3	Bulan	
4	Mata	
5	Ayam	
6	Badut	
7	Sandal	
8	Gajah	
9	Piring	
10	Buku	
11	Nanas	
12	Roti	
13	Semut	
14	Cecak	
15	Ember	
16	Pisang	
17	Jari	
18	Bayi	
19	Buah	
20	Bapak	
21	Dasi	
22	Bunga	
23	Anak	
24	Kapal	
25	Rambut	
26	Payung	
27	Lilin	
28	Burung	
29	Topi	
30	Gelas	
31	Telur	
32	Bebek	
33	Ikan	
34	Kursi	
35	Anjing	
36	Lampu	
37	Tangan	
38	Guling	
39	Sapi	
40	Gayung	
41	Garpu	
42	Kucing	
43	Coklat	
44	Tikus	
45	Bola	
46	Kambing	
47	Hidung	
48	Gunting	
49	Rumah	
50	Nasi	

Ear: **Right / Left or Binaural**

Unaided / Aided

Initial Presentation Level:

Date: / /

Name:

Age:

Audiologist:

Predicted SRT (in dBHL):
10.5 + (0.5 x 3 FAHL) ± 10

Mean:

N:

SE: SD/√N: ________
Should be < 2.5

Test Form A2

Item no.	Words	Excursions
1	Mata	
2	Bayi	
3	Bunga	
4	Bantal	
5	Kapal	
6	Rambut	
7	Nanas	
8	Gajah	
9	Lilin	
10	Burung	
11	Sandal	
12	Roti	
13	Semut	
14	Gelas	
15	Bebek	
16	Pisang	
17	Jari	
18	Kaki	
19	Buah	
20	Anak	
21	Dasi	
22	Rumah	
23	Ayam	
24	Bapak	
25	Garpu	
26	Gayung	
27	Piring	
28	Buku	
29	Topi	
30	Cecak	
31	Telur	
32	Ember	
33	Ikan	
34	Kursi	
35	Kambing	
36	Badut	
37	Tangan	
38	Gunting	
39	Anjing	
40	Payung	
41	Lampu	
42	Guling	
43	Bola	
44	Hidung	
45	Coklat	
46	Nasi	
47	Tikus	
48	Kucing	
49	Bulan	
50	Sapi	

Ear: **Right / Left or Binaural**

Unaided / Aided

Initial Presentation Level:

Date: / /

Name:

Age:

Audiologist:

Predicted SRT (in dBHL):
10.5 + (0.5 x 3 FAHL) ± 10

Mean:

N:

SE: SD/√N: ________
Should be < 2.5

Test Form B3

Item no.	Words	Excursions																	
1	Lilin																		
2	Topi																		
3	Kursi																		
4	Badut																		
5	Roti																		
6	Nasi																		
7	Ember																		
8	Bola																		
9	Anak																		
10	Bantal																		
11	Rambut																		
12	Ayam																		
13	Nanas																		
14	Ikan																		
15	Gelas																		
16	Gajah																		
17	Dasi																		
18	Rumah																		
19	Bulan																		
20	Lampu																		
21	Guling																		
22	Telur																		
23	Bapak																		
24	Bebek																		
25	Sapi																		
26	Bayi																		
27	Pisang																		
28	Piring																		
29	Payung																		
30	Bunga																		
31	Tikus																		
32	Mata																		
33	Tangan																		
34	Gunting																		
35	Coklat																		
36	Buku																		
37	Anjing																		
38	Buah																		
39	Burung																		
40	Cecak																		
41	Garpu																		
42	Kucing																		
43	Kapal																		
44	Gayung																		
45	Jari																		
46	Kaki																		
47	Semut																		
48	Hidung																		
49	Sandal																		
50	Kambing																		

Ear: **Right / Left**
or Binaural
Unaided / Aided
Initial Presentation Level:

Date: / /
Name:
Age:
Audiologist:

Predicted SRT (in dBHL):
10.5 + (0.5 x 3 FAHL) ± 10

Mean:

N:

SE: SD/√N: ________
Should be < 2.5

Test Form B4

Item no.	Words	Excursions																	
1	Lilin																		
2	Roti																		
3	Gunting																		
4	Garpu																		
5	Topi																		
6	Anjing																		
7	Bebek																		
8	Coklat																		
9	Kapal																		
10	Nanas																		
11	Badut																		
12	Anak																		
13	Mata																		
14	Pisang																		
15	Gelas																		
16	Tangan																		
17	Kaki																		
18	Rumah																		
19	Bunga																		
20	Lampu																		
21	Kucing																		
22	Semut																		
23	Ayam																		
24	Ember																		
25	Sapi																		
26	Bayi																		
27	Ikan																		
28	Piring																		
29	Gayung																		
30	Buah																		
31	Hidung																		
32	Bantal																		
33	Gajah																		
34	Kursi																		
35	Bola																		
36	Buku																		
37	Kambing																		
38	Bulan																		
39	Burung																		
40	Cecak																		
41	Rambut																		
42	Guling																		
43	Bapak																		
44	Payung																		
45	Dasi																		
46	Jari																		
47	Telur																		
48	Tikus																		
49	Sandal																		
50	Nasi																		

Ear: **Right / Left**

or Binaural

Unaided / Aided

Initial Presentation Level:

Date: / /

Name:

Age:

Audiologist:

Predicted SRT (in dBHL):
10.5 + (0.5 x 3 FAHL) ± 10

Mean:

N:

SE: SD/√N: ________
Should be < 2.5

APPENDIX G:
CLASSIFICATION OF INDONESIAN BISYLLABIC WORDS BASED ON CONTINUITY OF WAVEFORM ENVELOPE (SYLLABLE PULSES)

No.	Words	Single Pulse	Time of length of the word (second)
1	Lilin		0.652
2	Ember		0.777
3	Bola		0.866
4	Anak		0.519
5	Rambut		0.604

No.	Words	Single Pulse	Time of length of the word (second)
6	Ayam		0.700
7	Nanas		0.797
8	Gelas		0.809
9	Rumah		0.711
10	Bulan		0.640
11	Guling		0.773

No.	Words	Single Pulse	Time of length of the word (second)
12	Telur		0.628
13	Bayi		0.628
14	Piring		0.665
15	Payung		0.759
16	Bunga		0.742
17	Tangan		0.700

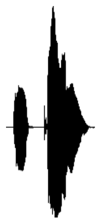
No.	Words	Single Pulse	Time of length of the word (second)
18	Buah		0.628
19	Burung		0.736
20	Kera		0.700
21	Gayung		0.833
22	Jari		0.724
23	Semut		0.688

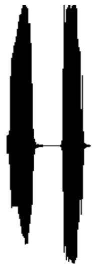
No.	Words	Single Pulse	Time of length of the word (second)
24	Sandal		0.869
25	Nenek		0.471
26	Tomat		0.519
27	Elang		0.676
28	Ular		0.640
29	Penyu		0.652

No.	Words	Single Pulse	Time of length of the word (second)
30	Jeruk		0.688
31	Mulut		0.495
32	Gunung		0.821
33	Kambing		0.797
34	Ceret		0.484
35	Domba		0.809

No.	Words	Single Pulse	Time of length of the word (second)
36	Anjing		0.773
37	Anggur		0.761

No.	Words	Two Pulses	Gap/ interval time	Time
1	Topi		159 ms	0.640
2	Kursi (dip)		259 ms	0.785
3	Roti		189 ms	0.550

No.	Words	Two Pulses	Gap/ interval time	Time
4	Nasi		205 ms	0.711
5	Bantal		93 ms	0.797
6	Ikan		120 ms	0.700
7	Gajah		154 ms	0.604
8	Dasi		165 ms	0.748
9	Lampu		104 ms	0.680

No.	Words	Two Pulses	Gap/ interval time	Time
10	Bapak		156 ms	0.507
11	Bebek		125 ms	0.567
12	Sapi		156 ms	0.833
13	Pisang		135 ms	0.785
14	Tikus		156 ms	0.845
15	Mata		145 ms	0.748

No.	Words	Two Pulses	Gap/ interval time	Time
16	Gunting		78 ms	0.833
17	Coklat		172 ms	0.616
18	Buku		142 ms	0.676
19	Garpu		118 ms	0.736
20	Kucing		89 ms before the bust of C or 177 ms	0.773
21	Katak		135 ms	0.869

No.	Words	Two Pulses	Gap/ interval time	Time
22	Kaki		112 ms	0.785
23	Hidung		88 ms	0.785
24	Jagung		106 ms	0.748
25	Sisir (dip)		183 ms	0.881
26	Gigi		112 ms	0.761
27	Poci		59 ms before the burst of C or 206 ms	0.748

No.	Words	Two Pulses	Gap/ interval time	Time
28	Kunci		41 ms before the burst of C or 148 ms	0.881
29	Kijang		106 ms	0.724
30	Lidah		106 ms	0.676
31	Cecak		189 ms	0.507
32	Kuda		106 ms	0.724
33	Sapu		165 ms	0.869

No.	Words	Two Pulses	Gap/ interval time	Time
34	Babi		130 ms	0.733
35	Pintu		84 ms	0.638
36	Ibu		141 ms	0.688
37	Kapal		156 ms	0.731

APPENDIX H:
ACCOMPANYING CD-ROM FOR INDO-SPASP WORDS

Appendix H consists of headings for sections of an accompanying CD-ROM named INDO-SPASP. The headings for the parts of this CD-ROM are:

1. **Individual (DU)** – This is a folder of Individual unfiltered words which have two pulses. These words are gigi, kursi, pintu, sapi and topi.

2. **Individual (QU)** – This is a folder of individual unfiltered words which have four pulses. These words are buku-buku, kapal-kapal, kuda-kuda, sapu-sapu and topi-topi.

3. **Individual (SU)** – This is a folder of individual unfiltered words which have single pulse. These words are bola, gelas, nanas, nenek and rumah.

4. **Individual double filtered (DF)** – This is a folder that contains the rhythmic pattern of the individual words which have two pulses. These words are gigi, kursi, pintu, sapi and topi.

5. **Individual quadruple filtered (QF)** – This is a folder that contains the rhythmic pattern of the individual words which have four pulses. These words are buku-buku, kapal-kapal, kuda-kuda, sapu-sapu and topi-topi.

6. **Individual single filtered (SF)** – This is a folder that contains the rhythmic pattern of the individual words which have single pulse. These words are bola, gelas, nanas, nenek and rumah.

Note:

1. Unfiltered – Before the words in each list were put through the HELOS (hearing loss simulator).

2. Filtered – After the words in each list were put through the HELOS.

APPENDIX I:
PHOTOGRAPHS OF SCHOOL

The partition and walls of the school were made from webbing bamboo.

APPENDIX J: INFORMATION LETTERS AND CONSENT FORMS

Adult Participants

Date Participant's copy

Dear Participant,

Name of the project: **The development of speech tests for children in the Indonesian language or Bahasa Indonesia (BI)**

I would like to invite you to participate in a research project that is being conducted as part fulfillment of the Doctorate Degree in Audiology under the supervision of Prof. Philip Newall and Dr. Robert Mannell. The aim of this study is to develop a speech test for children in the Indonesian language or "Bahasa Indonesia (BI)". To do this I will need to assess the speech test item that is being developed. This will involve testing the child's ability to hear and understand spoken language using the speech test item.

At the initial stage of this research, the reliability of the speech test item needs to be assessed with adult participants first before testing it with children. Therefore, your participation in this research would be highly appreciated.

During this study, your hearing will be tested using standard clinical hearing tests which involve responding to low level sounds presented by earphones and other technique which involve minimal co-operation. You will then be asked to listen to a number of words in the Indonesian language using headphones. You will be asked to repeat the word you hear.

The study will be carried out at the Royal Institute for Deaf and Blind Children, North Rocks, NSW. Should you decide to participate in the study, any information or personal details gathered in the course of the study are confidential. No individual will be identified in any publication of the results. I would also be happy to share the results of your hearing assessment.

If you would like to be in this study, please sign the consent form at the end of this letter. You are free to withdraw your consent at any time, without having to give a reason and without adverse consequence. The data that we collect from you may be used by other researchers, but that data will not reveal your details in any way.

Please also note that the ethical aspects of this study have been approved by the Macquarie University Ethics Review Committee (Human Subjects). If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Ethic Committee through its Secretary (telephone 9850 7854). Any complaint you make will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you have any question about the study, please feel free to contact me and I will be happy to discuss the study with you.

Yours Sincerely,

Dr. Dahlia E. Sartika	Robert Mannell, PhD	Prof. Philip Newall
PhD Candidate	Supervisor	Supervisor
Ph: 612 9636 3403	Ph: 612 9850 8771	Ph: 612 9850 8779

Consent Form

I have read (or where appropriate, have read to me) and understand the information above, and any questions I have asked, have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw at any time. I have been given a copy of this form to keep

Name of participant: … … … … … … … … … … … … … … … … (Block letters)

Signature of participant: … … … … … … … … … … … … … … … Date: … … … …

Name of investigator: … … … … … … … … … … … … … … … … (Block letters)

Signature of investigator: … … … … … … … … … … … … … … … Date: … … … …

Consent Letter in English for Children

Date Participant's copy

Dear Parent,

Name of the project: **The development of a speech test for children in Indonesian language or Bahasa Indonesia (BI)**

I would like to have your child participate in a research project that is being conducted as part fulfillment of the Doctorate Degree in audiology under the supervision of Prof. Philip Newall and Dr. Robert Mannell. The aim of this study is to develop a speech test for children in the Indonesian language or Bahasa Indonesia (BI). To do this I will need to assess the speech test item that is being developed. This will involve testing the child's ability to hear and understand spoken language using the speech test item.

Your child's hearing will be tested using standard clinical hearing tests which involve responding to low level sounds presented by earphones and other technique which involve minimal co-operation from the child. Your child will then be asked to listen to a number of words in the Indonesian language using headphones. They will be asked to point to one of a set of pictures corresponding to the word presented.

The actual assessment will be carried out at the Audiology clinic, Macquarie University and in Jakarta, Indonesia at PT Kasoem Hearing Centre. Should you decide to allow your child to be in the study, any information or personal details gathered in the course of the study are confidential. No individual will be identified in any publication of the results. I would also be happy to share the results of your child's assessment with you.

If you would like your child to be in this study, please sign the consent form at the end of this letter. You are free to withdraw. You are free to withdraw your consent at any time, without having to give a reason and without adverse consequence. The data that we collect from your child may be used by other researchers, but that data won't identify your child in any way.

Please also note that the ethical aspects of this study have been approved by the Macquarie University Ethics Review Committee (Human Subjects). If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the Ethic Committee through its Secretary (Phone 9850 7854). Any complaint you make will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you have any question about the study, please feel free to contact me and I will be happy to discuss the study with you.

Yours sincerely,

Dahlia Eka Sartika	Robert Mannell, PhD	Prof. Philip Newall
Postgraduate Student	Supervisor	Supervisor
Ph: 612 9636 3403	Ph: 612 9850 8771	Ph: 612 9850 8779

Consent Form

I (the parent or guardian of the participant) have read (or where appropriate, have read to me) and understand the information above, and any questions I have asked, have been answered to my satisfaction. I agree to have my child participate in this research, knowing that I can withdraw at any time. I have been given a copy of this form to keep.

Name of the participant's parent or guardian: … … … … … … … … … (block letters)

Signature of the participant's parent or guardian:… … … … … … … … Date: … … …

Investigator's Name:… … … … … … … … … … … … … … … … … (block letters)

Investigator's Signature:… … … … … … … … … … … … … … … …Date: … … …

Consent Letter in Indonesian for Children

Tanggal

Kepada yth,
Bapak/Ibu

Judul Proyek: *Pengembangan Tes Pendengaran Anak Dengan Tutur Kata Bahasa Indonesia*

Bersama ini saya ingin mengajak putra/putri anda untuk berpartisipasi dalam proyek penelitian yang akan saya lakukan sebagai syarat gelar Doktor (S3) dalam bidang audiology (ilmu pendengaran).

Tujuan penelitian ini untuk mengembangkan tes pendengaran anak dengan tutur kata bahasa Indonesia. Untuk hal tersebut saya perlu menguji tes pendengaran dengan tutur kata yang sedang dikembangkan tersebut. Pengujian tersebut akan melibatkan kemampuan anak mendengar dan mengerti bahasa Indonesia dengan tutur kata yang menggunakan tes pendengaran tersebut.

Pendengaran anak anda akan diuji menggunakan tes-tes pendengaran klinik standard, dimana anak akan merespons nada-nada pelan melalui headphones dan tehnik-tehnik lainnya yang hanya memerlukan kerja sama minimal dari anak. Kemudian anak anda akan diminta untuk mendengarkan sejumlah kata-kata melalui headphones. Mereka diminta untuk menunjukkan gambar yang sesuai dengan kata yang didengar.

Pengujian akan dilaksanakan di klinik audilogy-PT Kasoem, Jakarta. Sekiranya anak anda untuk turut berpartisipasi dalam penelitian ini, maka semua informasi dan data pribadi yang tercatat dalam penilitian ini akan dijaga kerahasiaannya. Begitu juga dalam publikasi hasil penelitian, seluruh identitas individu tetap akan dirahasiakan.

Jika anda mengizinkan anak anda menjadi peserta, maka kami memerlukan tanda tangan anda pada formulir dibawah ini. Tetapi bila anda ingin mengundurkan diri, anda bebas melakukannya setiap waktu tanpa harus ada alasan atau izin khusus dan tidak akan berkonsekuensi buruk. Data penelitian yang kami dapatkan dari anak anda, dapat digunakan oleh peneliti lain, tapi dalam data tersebut tidak akan ditemukan identitas peserta.

Jika anda ingin mendapatkan informasi tentang hasil penelitian atau apabila anda mempunyai keluhan, keberetan tentang aspek-aspek ethis keikut-sertaan anda dalam penelitian ini, anda dapat menghubungi saya melalui PT Kasoem. Setiap keluhan anda akan kami tanggapi dengan penuh kesungguhan dan kerahasiaan dan hasilnya akan kami informasikan.

Bila anda mempunyai pertanyaan tentang penelitian ini, silahkan hubungi saya dan saya akan senang berdiskusi dengan anda.

Dengan Hormat,

Dr. Dahlia Sartika
(MD, PGDipAud, Maud/CCP)
Telepon: +628881328950
Email: dahliasartika@yahoo.com.au

Formulir Persetujuan

Saya (orang tua atau wali dari peserta) telah membaca (atau dimana perlu, telah dibacakan kepada saya) dan mengerti keterangan diatas. Semua pertanyaan yang saya ajukan telah dijawab dengan memuaskan.

Saya menyetujui anak saya turut berpartisipasi dalam penelitian ini dan saya mengerti bahwa saya dapat mengundurkan diri setiap waktu. Saya juga telah mendapatkan fotokopi dari formulir ini.

Nama orang tua/wali peserta……………………………………………………………………………...
……………………………….. (tulis dengan huruf balok)

Tanda tangan orang tua/wali peserta………………………………………….....Tanggal…………………..

Nama peneliti……………………………………………………………………..(tulis dengan huruf balok)

Tanda tangan peneliti……………………………………………………………......Tanggal…………………….

Appendix K of this thesis ("Final ethics approval letter") has been removed as it may contain sensitive/confidential content