

Ellen Saxlund

'Make Monkey hide!'

A descriptive account of the NRDLS and the Norwegian adaptation

Master's thesis in English Linguistics
Supervisor: Mila Dimitrova Vulchanova
May 2019



Ellen Saxlund

'Make Monkey hide!'

A descriptive account of the NRDLS and the
Norwegian adaptation

Master's thesis in English Linguistics
Supervisor: Mila Dimitrova Vulchanova
May 2019

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

The cover picture depicts Monkey, one of the main characters of the New Reynell Developmental Language Scales, looking at item 26 from comprehension section (Cii).

The quote in the title is taken from item 25 of the comprehension part of the NRDLS, where the child is asked by the test administrator to make Monkey perform certain tasks.

Picture © Ellen Saxlund

The New Reynell Developmental Language Scales TM are the property of GL assessment and Hogrefe, all rights reserved.

Abstract

The New Reynell Developmental Language Scales (NRDLS) is a test instrument intended for use in clinical contexts, primarily for detecting possible atypical language development in children. It is based on several previous versions and revisions of the test, which was published originally in the UK in 1969. In 2017-18, the NRDLS was adapted into Norwegian and standardised for publication and use with a Norwegian-speaking population, for the first time since 1985. The current Norwegian NRDLS was first piloted on a total of 50 children, and for the standardisation and collection of normative data, 242 children between 3:00 and 7:10 years were tested. This thesis aims to be a descriptive account of the NRDLS and its theoretical basis, test adaptation, as well as the process of adaptation and standardisation of the Norwegian NRDLS. When adapting an existing test instrument into a different language, there are several considerations that need to be accounted for. For the NRDLS, the most important was item equivalence and construct equivalence, as well as familiarity with milestones in language development in children. This thesis also discusses some findings from the standardisation process, such as the subtle gender differences in the performance of Norwegian children on the NRDLS, the limited participation of children younger than three years, and possible uses of the NRDLS other than tracking language development in children.

Sammendrag

New Reynell Developmental Language Scales (NRDLS) er testverktøy som er tiltenkt bruk i kliniske kontekster, hovedsakelig for å avdekke atypisk språkutvikling hos barn. NRDLS er basert på flere tidligere versjoner og revisjoner av testen, som originalt ble utgitt i Storbritannia i 1969. I 2017-18 ble NRDLS tilpasset og oversatt til norsk og standardisert for bruk på barn med norsk som morsmål, for første gang siden 1985. Den nåværende norske versjonen av NRDLS ble pilottestet på 50 barn med norsk som morsmål. For standardiseringen ble 242 barn mellom 3:00 og 7:10 år testet. Denne oppgaven har til hensikt å være en beskrivelse av NRDLS og dens teoretiske bakgrunn, generelt om tilpasning og oversettelse av testverktøy, i tillegg til tilpasningen og standardiseringen av den norske NRDLS. Man må ta hensyn til flere ting når et testverktøy skal oversettes og tilpasses et annet språk. I NRDLS' tilfelle var det viktigste hensynet konseptuell ekvivalens, i tillegg til kunnskap om milepæler i språkutviklingen hos barn. Denne oppgaven diskuterer også noen funn fra standardiseringsprosessen, som blant annet subtile kjønnsforskjeller i resultatene hos barn med norsk som morsmål, samt den begrensede deltakelsen av barn under tre år. I tillegg diskuteres mulig bruk av NRDLS utover å kartlegge språkutviklingen hos barn.

Foreword

In October 2017, my supervisor, Mila Dimitrova Vulchanova, inquired for students to join a project where we were asked to do some field work and data collection. I decided to join, without knowing exactly what I had signed up for. The project in question was the standardisation and norm development of the Norwegian version of the New Reynell Developmental Language Scales, or the NRDLs, a test instrument used mainly for detecting atypical language development in children. Together with my supervisor, Randi Alice Nilsen, and Valentin Vulchanov, myself and five other students took part in the piloting of the NRDLs during the last months of 2017 and beginning of 2018, where a total of 50 normally developing children were tested. During the last half of 2018, we tested nearly 250 children between the ages of 2 and 7 years for the standardisation and norm development of the NRDLs.

Firstly, I would like to thank my supervisor, Mila Vulchanova, for allowing me to take part in this project. I am immensely grateful to have been a part of it, and it has been an invaluable experience, both regarding language research and data collection, but also for my future work as a teacher. Moreover, thank you for all the helpful comments and tips for this thesis. I would also like to give a big thank you to Stefan Williams, my closest colleague with whom I conducted most of the testing sessions.

I would also like to thank Arunima Vijay for letting me contribute with the testing for her project on ASD and mental verbs. Yet again, the experience has been invaluable, and I am truly grateful to have taken part in the project.

A big thank you should be given to all the schools, kindergartens, and parents who let us test their children. Thank you for being so patient and giving us your time. I can honestly say that I have met so many clever, lively and amazing children, and it has truly been a great joy to test every single one of them.

My deepest thanks go out to all of my close friends and family, who are always there to support me and cheer me on. Especially Emilie Haudecoeur Davidsen, Greatania Juardi and Hanna Jarstø Ervik each deserve an honourable mention for all the hours you have spent with me, and for always bringing joy into my life.

And lastly, a big thank you to my boys.

Trondheim, May 2019

Ellen Saxlund

Abbreviations

ASD = Autism spectrum disorder

ASQ = The Ages and Stages Questionnaire

L1 = First language(s)

L2 = Second language(s)

MacArthur-Bates CDI = MacArthur-Bates Communicative Development Inventory

NRDLS = New Reynell Developmental Language Scales

PPTV = The Peabody Picture Vocabulary Test

SES = Socio-economic status

(S)LI = (Specific) language impairment

TROG-2 = Test for Reception of Grammar

Contents

Foreword	vii
Abbreviations.....	ix
1.0 Introduction.....	13
1.1 Structure and composition.....	13
2.0 Standardised tests	15
2.1 What are standardised and norm-referenced tests?	15
2.2 Different types of clinical (language) tests	15
2.3 The utility of standardised tests.....	17
2.3.1 The utility of the NRDLS	18
3.0 The New Reynell Developmental Language Scales	21
3.1 The structure, content and administration of the NRDLS	21
3.1.1 The development of the NRDLS.....	23
3.2 Theoretical background.....	24
3.2.1 Contact-establishing tasks	25
3.2.2 Single words.....	25
3.2.3 Locatives.....	26
3.2.4 Verbs.....	26
3.2.5 Pronouns	28
3.2.6 Complex sentence structures.....	28
3.2.7 Inferences.....	29
3.2.8 Grammaticality judgement	29
4.0 Adapting clinical instruments into different languages.....	31
4.1 Generally on test adaptation.....	31
4.1.1 Guidelines for test development.....	33
4.1.2 Guidelines for test adaptation	34
4.2 Developing clinical tools into different languages	37
4.2.1 TROG-2	37
4.2.2 ASQ.....	38
5.0 The Norwegian NRDLS	41
5.1 The standardisation process	41
5.1.1 Recruiting informants.....	41
5.1.2 Item number 25 and object permanence	43
5.1.3 Difficulties with children under 3:00 years	44

5.2 Gender differences	46
5.2.1 Gender differences in ten different language communities	47
5.2.2 Case study on gender and diminutives	49
5.3 The adaptation from Norwegian into English	51
5.4 The possible uses of the NRDLS.....	52
5.4.1 NRDLS and socio-economic factors.....	53
5.4.2 Using the Multilingual Toolkit or the standard NRDLS?	54
5.4.3 Other possible uses and future research	55
6.0 Conclusion	57
7.0 Literature.....	59
Appendices	
Appendix A	
Appendix B	
Appendix C	
Appendix D	
Table 1.0	
Appendix E	
Relevance for the Teacher Education	

1.0 Introduction

In general, standardised tests for clinical, research and educational purposes have been in use for a long time, and adaptation of these into other languages allow for a greater number of people to use existing and well-tested instruments for clinical and research purposes.

However, there are several considerations that must be taken into account when adapting such test instruments into other languages, such as language use and which constructs the test measures.

One such clinical assessment instrument that has been recently adapted into Norwegian is the New Reynell Developmental Language Scales, or the NRDLs. The NRDLs is a clinical tool used to assess language development in children, and to detect any possible atypical language development. First published in 1969 in the UK as the Reynell Developmental Language Scales, the language scales have undergone a series of revisions and republications, leading up to the current version, which was published for use in English language contexts in 2011. Per early 2019, the language scales and the manual have only been published once before in Norwegian, and this was in 1985 (Hagtvet & Lillestølen, 1985). Thus, the need for a revised Norwegian version of the language scales appeared to be imminent. Through the adaptation and standardisation of the Norwegian NRDLs, 50 Norwegian children were pilot-tested, and nearly 250 children were tested for the development of the Norwegian norms.

This thesis aims to be a descriptive account of the NRDLs and its possible uses, the theoretical background for its parts, as well as the process of adaptation and standardisation of the Norwegian NRDLs. The thesis will also account for some considerations regarding development and adaptation of standardised language tests for primary use in clinical contexts.

1.1 Structure and composition

The thesis is divided into four main chapters. Chapter 2 on standardised testing describes what standardisation means, the utility of standardised tests, as well as some different types of standardised language tests. Chapter 3 goes into more detail about the development of the NRDLs, as well as providing a theoretical background for each part of the test. In chapter 4 the process of adapting clinical instruments into different languages is described, as well as two accounts of clinical test instruments that have been adapted into other languages. Lastly, chapter 5 discusses the process of the adaptation and standardisation of the Norwegian

NRDLS, as well as detailing some possible suggestions for uses of the NRDLS, as well as possible future research.

2.0 Standardised tests

2.1 What are standardised and norm-referenced tests?

Standardisation of tests essentially refers to a process wherein test instruments are developed, piloted and revised. Through this process, a standard for test administration is developed. This way, instructions and standards for test administration will be the same regardless of who administers the test and in what setting the test is administered, ensuring that there is inner consistency and high reliability (Shawer, 2010). Commonly, a standardised test is first pilot tested on a small-scale population before the actual testing and data collection takes place. This is an important step in the standardisation of test instruments, as it is a way of discovering and evaluating any issues with the design, costs, or difficulties that may arise during testing, such as difficulties with the administration, certain test items and the wording of them or similar. By piloting a test, it is possible to adjust or change certain items or aspects of the instrument accordingly, in order to ensure that the instrument will work on a larger scale, and that all test items are feasible (Thabane et al., 2010). If any test items or other aspects of the instrument is not working optimally, it will then be possible to adjust this before the larger, main project is carried out.

Another important step in test development is the norm-referencing of a test instrument. The process of norm-referencing generally involves sampling data from a large number of participants in a certain part of a population, depending on what the test measures (Cronbach, 1990). The norm is then estimated by analysing the sampled data and test scores from the selected population (Kirk & Vigeland, 2014). What makes norm-referenced tests effective is partly that they provide researchers and clinicians the possibility to directly compare the score of one individual to the performance of a number of other individuals of the same age group, who were tested under the same conditions (Kaufman & Kaufman, 2004). In that way, any individual in society can be tested using the standardised test, and by comparing their score to the norm-reference for the particular age bracket it is possible to say something about their performance, if they are average, above average, or possibly in need of intervention.

2.2 Different types of clinical (language) tests

When talking about standardised language tests for clinical purposes, they come in a wide variety, depending on what is tested. There exists wide varieties of standardised language and

psychological test instruments for children, which may be useful in both clinical and scientific settings. Some of these, in addition to the NRDLs, are: the Ages and Stages Questionnaire (ASQ) (Bricker & Squires, 1999), The Test of Word Finding (German, 2000), the Test for Reception of Grammar (TROG-2) (Bishop, 2003), the Test of Early Grammatical Impairment (Rice & Wexler, 2001), the Early Registration of Language Development (TRAS) (Espenakk, 2003), and the Peabody Picture Vocabulary Test (PPVT-4) (Dunn & Dunn, 2007).

There are so-called comprehensive language tests, that test a wide variety of both receptive and productive language. One such example is the NRDLs, which tests both language comprehension and language production. This is thus an example of a complete comprehensive language test. There are also comprehensive tests for just one aspect of language, such as receptive language. TROG-2 is an example of this, as it tests only receptive grammar. For a more detailed discussion of TROG-2, see section 4.2.1 There are also numerous language tests that test more specific domains. For instance, one test often used with children with ASD (Autism spectrum disorders), is the TOPL-2 (Test of Pragmatic Language), and it only tests pragmatic language specifically (Phelps-Terasaki & Phelps-Gunn, 2007).

There are also differences in how tests are administered. For instance, the NRDLs and TROG-2 are examples of tests where a test administrator conducts the test directly together with the individual being tested and is thereby able to directly observe and record their responses. ASQ and the MacArthur-Bates Communicative Development Inventories (CDI) are examples of parent-reported questionnaires, where parents are asked to observe their child over a period of time and then record the behaviour in the questionnaire (Squires, Bricker & Potter, 1999; Fenson et al., 1993). There are both advantages and disadvantages to both types of administration, as parent-administered questionnaires allow for a wider range of questions, and the child can be observed for a longer period of time. However, as the data collected by the researcher or clinician is only based on second-hand accounts by parents, it is debatable how reliable the data is. As Tsai, McClelland, Pratt and Squires (2006) note in their experiment when distributing the ASQ to both parents and teachers, the parents generally tended to give their own children a higher score than the teachers did (Tsai et al., 2006). For a more detailed description of the ASQ, see section 4.2.2. Nevertheless, different types of test instrument should be used depending on the need, the purpose and the circumstances of the testing.

2.3 The utility of standardised tests

A key question in any research or clinical contexts involving the use of standardised tests is why these types of tests are useful. Especially concerning use in scientific contexts, standardised tests may be limiting to the researcher in answering the intended research question, and especially in cases where test instruments have been adapted to a different linguistic or cultural context. However, for clinical use standardised tests have proved to be quite useful tools for speech therapists, paediatricians, psychologists, teachers and anyone working with children and cognitive and language development (Bishop, 2003).

Standardised and norm-referenced tests for clinical use are in general effective for identifying individuals in need of early intervention. Moreover, they are instruments of justifying this early intervention to parents and caregivers (Edwards, Letts & Sinka, 2011). A test such as NRDLS or TROG-2 may also be used to chart the progress of an individual after intervention. A child can be tested by using a standardised test, it may be discovered that the child is in need of some type of intervention, and these measures are then implemented. After some time has passed, the child can be tested using the same standardised test again. This way, it is possible to measure whether the interventive measures were successful or not (Fava et al., 2011).

A last concern remains, however, as Brislin (1986) notes, that there seems to be a common belief that existing test instruments are always sound, and function as some sort of 'golden standard' for research instruments. This may lead to a false sense of security, as existing test instruments need not be appropriate in all contexts, especially in scientific contexts. Furthermore, as Brislin (1986) notes, it may be as useful for a researcher to design and develop their own tools rather than relying completely on existing instruments, as the researcher has the opportunity to design their research in accord with their specific research question and needs. There is also a question whether standardised tests are able to measure all subtle language deficits, such as subtle deficits that may be found in children of a low socio-economic status. Fazio, Naremore and Connell (1996) note that children with a lower socio-economic status may have language deficits that are not completely captured through the use of standardised tests. It is important to note that standardised tests only have a limited use, and a standardised test can essentially only provide background information of an individual. If a specific research question is to be tested, it would in most cases be more fruitful to design a completely new experiment.

Nevertheless, standardised test instruments have been and continue to be a useful tool for use in clinical contexts, for identifying individuals that may need early intervention, and to have a means of justifying this precise intervention to parents of a child that appears to experience some developmental delay.

2.3.1 The utility of the NRDLs

As mentioned, there are several uses of standardised tests in both clinical and experimental contexts. Regarding the NRDLs, there are several reasons why it might be a particularly useful clinical tool, and why the adaptation of the NRDLs into Norwegian was necessary.

Firstly, the results from the NRDLs are highly transparent, and give an immediate result after the testing is finished and the responses are counted and summarised. The results provide both an image of the current state of the language development of an individual and if there are any apparent language delays. Additionally, it is also possible to use a qualitative analysis of the results to determine if there is any need of further examination of any specific areas that the NRDLs tests, such as verb endings, complex sentences or vocabulary (Edwards et al., 2011). As each of the sections in the NRDLs contains a relatively low number of items, it is not possible to obtain a definitive indication of delay in language development in any of the areas that the test examines. NRDLs is not intended to be used as a tool for final diagnostic of language impairments, but can be used as a diagnostic tool to justify further examination. However, the NRDLs can be a good starting point for examination of any difficulties or problem areas that may be present in an individual, and since it covers many different types of competences, there is a possibility to use the NRDLs to discover a large variety of language difficulties or delays an individual may experience (Vulchanova, Vulchanov & Nilsen, 2019). Moreover, as Letts, Edwards, Schaefer, and Sinka (2014) point out, the NRDLs may be more appropriate for monitoring general language development over time, rather than measuring specific language problems in individuals at a given point in time.

In addition, in younger children, especially those younger than 3 years of age, environmental factors are less of a prominent influence on general developmental trends (Janson & Squires, 2004). Thus, the NRDLs may be especially useful in mapping children with developmental delays before they reach the age of around 3:00, as factors such as environment and socio-economic background of the parents will not necessarily impact the general developmental patterns of the child (Janson & Squires, 2004). It should be noted that for the standardisation

of the Norwegian NRDLs, norm-data for children under 3:00 was ultimately not collected, see section 5.1.3 for a further discussion of this.

Furthermore, the NRDLs may be more useful than other tools for detecting language or general developmental delays in children which are commonly used in Norway, such as TRAS, SATS, TROG-2 and SPRÅK 4. The reason being that the NRDLs covers a larger age range, and it is also a complete comprehensive language test, thus testing a broader scope of competences, as it tests both language comprehension and language production (Vulchanova et al., 2019). Lastly, the NRDLs may have the potential to be adapted slightly to be used for testing people with cochlear implants and older immigrant adults in oral language competence, see section 5.4.3 for a discussion on this.

3.0 The New Reynell Developmental Language Scales

3.1 The structure, content and administration of the NRDLs

The first version of the Reynell language scales was developed in the UK in the 1960s, and subsequently published in 1969, under the name of the Reynell Developmental Language Scales. It was developed to evaluate language in both clinical and scientific contexts (Edwards et al., 2011). It has since been revised and republished several times, and the NRDLs is the fourth version of the language scales.

The NRDLs is divided into two main parts, one for language comprehension, and one for language production. The comprehension part consists of 72 items, which are divided into 8 sections. It also has a warm-up task, intended for the youngest children, those who are suspected to be developmentally delayed, or shy children, as a form of contact-establishing task. Here, the child is asked to point to various body parts as named by the test administrator. The 8 sections in the comprehension part consist of the following types of tasks: (A) Selection of Objects, (B) Two Objects in Relation, (C) Intransitive Verbs, (D) Sentences with Transitive and Intransitive Verbs, (E) Verb Morphology and Verb Endings, (F) Pronouns, (G) Complex Sentences, and (H) Inferences.

The production part consists of 64 questions, which are divided into 7 sections. This part also has a warm-up task, where children are asked to name various body parts the test administrator points to. The 7 main sections in this part consists of the following type of tasks: (A) Naming Objects, (B) Two Objects in Relation, (C) Verbs, (D) Sentence Building, (E) Verb Morphology and Verb Endings, (F) Complex Sentences, and (G) Grammaticality Judgement. For further discussion of the theoretical background and what each section tests, see section 3.2 of this thesis.

Additionally, nearly every section has a parallel section in the other part of the test. For instance, section (E) of both the comprehension and the production parts test verb morphology and verb endings. The exceptions are sections (F) Pronouns and (H) Inferencing in the comprehension part, and section (G) Grammaticality judgement in the production part. The use of parallel sections in this way makes it possible to compare the results in two parallel sections against each other, and from there detect any possible developmental language impairments. There are some language impairments that will affect either language comprehension or language production but not both, such as verbal dyspraxia and aphasia

(Karmiloff & Karmiloff-Smith, 2001; Caramazza & Hillis, 1991). In this way, the NRDLs can be used to determine if there is a need for further testing for such language impairments in an individual.

The NRDLs consists of a picture book, with a series of characters and situations that should be familiar to young children, such as sliding down a slide, swimming, eating an apple, and reading a book (see appendix C, images 1 and 2). In addition, there are seven additional picture cards, for use in production section (Dii). There is a manual detailing the background for the development of the test, theoretical support for each section of the test, and detailed instructions on how to administer the test. There are test protocols for scoring and detailed instructions for each test item. In addition, the test also includes a series of objects that should be familiar to children, such as a ball, a rubber duck, a pencil, and three stuffed animals, which also frequently appear in the pictures presented to the children. See appendix C for images of the objects that the NRDLs contains. The items in use are objects commonly used as toys and should appear enjoyable to children in the target age group. Lastly, the NRDLs also includes the Multilingual Toolkit which is intended as a guide for administering the NRDLs to children that do not have English (or Norwegian, for the Norwegian NRDLs) as one of their L1s. For further discussion of the Multilingual Toolkit, see section 5.3.2.

When administering the test, the test administrator generally sits with the child at the corner of a table, where the child sits to the left of the test administrator. The test administrator has the objects to the right, either on a chair or a small table, obscured from the child's sight. The test leader reads out each test item from the test protocol, while presenting the objects or the images from the picture book to the child. Furthermore, for the administration of the test it is recommended that the tasks are presented to the children like a game, so that it appears less overwhelming and becomes a more enjoyable experience for the child (Vulchanova et al., 2019).

The NRDLs is also designed so that the items get progressively more difficult for each section. Thus, section (A) in both parts should be the easiest and all typically developing children from around 2:00 years and up should be able to successfully complete this section. Then the sections become progressively more difficult through to section (G) and (H), which only the oldest children should be able to answer. The idea is that the test should be stopped when a child fails two entire sections in a row (Edwards et al., 2011). When the child obtains a score of zero for two entire sections in a row in the comprehension part, the test

administrator should begin from section (A), alternatively from the warm-up section, of the production part. When the child fails two entire sections in a row in the production part, the testing should end. Although each section is more difficult than the next, some language features may develop parallelly in language acquisition, and thus the test administrator must ensure that they child truly is not able to perform adequately in the further sections (Edwards et al., 2011). For a more detailed discussion of the theoretical background for the linguistic aspects that are tested in each section, see section 3.2 of this thesis.

3.1.1 The development of the NRDLs

Through the numerous revisions of the language scales, as well as piloting of the NRDLs, each of the sections and items in the test have been well-tested and reviewed before publication. This includes the images and objects used in the test, which have also been tested for suitability for children. Through the process of the standardisation of the NRDLs in the UK, the NRDLs was first piloted on 151 girls and 150 boys between the ages of 1:06-7:06. For the standardisation, in 2009-10 a total of 1266 individuals in the UK between the ages of 2:00 and 7:05 were tested. A series of factors were taken into consideration when selecting the items for the final version of the NRDLs (Edwards et al., 2011).

First, there must be correlation between test scores in the sense that there is correlation between scores on single items and overall score, to ensure that any given item correlates well between children who obtained a high score and children who obtained a low score (Edwards et al., 2011).

Cronbach's alpha has been used to test for inner consistency. Cronbach's alpha was used to measure the reliability, and how removal of single items would affect reliability. Cronbach's alpha showed that only a few questions affected the reliability of the instrument if they were removed, which indicates good overall reliability (Edwards et al., 2011).

Another important consideration when selecting test items, was that the test should be able to measure a unique developmental level for each age bracket of six months. In the case of any test sections where two age brackets performed equally, one was removed. A previous section from RDLS-III on extended noun phrases was removed, as it was not able to measure a unique developmental effect in one unique age bracket (Edwards et al., 2011).

When piloting the NRDLs in the UK, it was found that a very limited number of the children in the age bracket 1:06-1:11 could respond sufficiently or wanted to cooperate, and thus this

age bracket was removed from the standardisation (Edwards et al., 2011). For the Norwegian NRDLS, the same phenomenon was found to occur for the age brackets 2:00-2:05 and 2:06-2:11 years. For a more in-depth discussion of this finding, see section 5.1.3 of this thesis.

3.2 Theoretical background

The NRDLS has been developed in order to discover specific linguistic developmental delays in children. Based on the type of sections where children make a significant amount of errors, upon analysis of the child's result, it is possible to gauge what type of delays the child might experience, and alternatively what measures need to be put in place. Each section of the NRDLS has specific target areas it tests, and as long as a child masters the intended target area in the specific section, they will pass the section despite any other errors they might make. There is much scientific and theoretical support for the NRDLS and its specific parts and intended use. The NRDLS is essentially developed based on expected linguistic developmental milestones, and each section tests specific grammatical features that are generally believed to occur at different points in the language development (Edwards et al., 2011).

For instance, Rice and Wexler (1996) have found that poor use of tense markers in English, that is, the singular -s in third person present singular and the regular -ed in past tense verbs, may be an indication of certain language impairments. Similarly, poor understanding and use of pronouns and reflexive pronouns may indicate the presence of an (S)LI (Lely & Stollwerck, 1997). Additionally, problems with complex sentences such as passive constructions, WH-questions and relative clauses may be indicative of language impairments (Letts et al., 2014). The NRDLS has sections for all of these competences, and difficulties with these outside of what can be expected according to the child's age bracket can provide indication that other test instruments may be utilised to examine these possible language impairments further. It is also recommended that the test administrator notes down if children need many of the items to be repeated, as this may be a sign of problems with working memory (Letts et al., 2014).

The trajectory of language development in children generally follows the same path across languages. It seems that in general, children across languages and cultures go through roughly the same stages of language acquisition in the same order (Karmiloff & Karmiloff-

Smith, 2001). Thus, in the Norwegian adaptation all of the sections are kept in the same order as the English version, and the content of each section remains largely the same.

3.2.1 Contact-establishing tasks

The first section of both parts of the NRDLs, which is the contact establishing tasks, are not actually part of the scored sections, and do not affect the results as a whole. However, they can be highly beneficial for children who are very young, who are shy, or who are suspected to have linguistic or cognitive difficulties. The main goal of this section is to grab the child's attention, create a relaxed and safe atmosphere and establish contact (Edwards et al., 2011). This may affect the child positively and ensure that their performance is optimal. According to Sheridan, Sharma and Frost (1997) children as young as 1:00 years may be able to participate in this type of activity, and thus it should be a good way of establishing contact and setting the tone for the testing for most children.

3.2.2 Single words

In section (A) of both the comprehension and production part, children's knowledge of single words is tested. The words in the NRDLs are presented in order of frequency in English (Dale & Fenson, 1996; Morrison, Hirsh & Duggan, 2003; Masterson, Druks & Gallienne, 2008) In the Norwegian adapted version, the equivalent words are used in the same order. There seems to be evidence for the fact that in general, nouns are acquired before verbs, regardless of language (Caselli et al., 1995). At least one notable exception is found in Korean, where the acquisition of nouns does not necessarily happen before verbs, or at least more verbs seem to be acquired at an earlier stage than what has been observed in other languages (Kim, McGregor & Thompson, 2000). Nevertheless, as the Norwegian and English languages largely have similar linguistic features, share a large portion of grammatical structures, as well as the fact that the cultural differences in how language is acquired and how children are raised is deemed to be relatively similar, it may be safe to assume that the acquisition of Norwegian nouns follow a relatively similar trajectory to the English acquisition.

According to Masterson, Druks and Gallienne (2008), nouns are acquired earlier than verbs, and thus the sections on nouns (section (A)) comes before the section on verbs (section (C)) in both parts of the test. In general, verbs tend to be more complex than nouns, as they generally do not correspond to any physical entities in the world, as opposed to nouns. Especially high-frequency nouns that are acquired early in the acquisition tend to correspond

directly to physical objects that are common in children's environment, and that parents tend to use. Moreover, according to Gleitman (1994), it is more difficult for children to generalise from one verb to another when acquiring verbs, as verbs tend to be more complex semantically, they denote more referents, and generally involve more complex situations than nouns. It is also more difficult to relate verbs to concrete entities in the physical world (Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005).

3.2.3 Locatives

One of the earlier sections, section (B) in both the comprehension and the production part, has to do with relating two objects, locatives, and semantic understanding. An example taken from the comprehension part is item number 15, where the test administrator asks the child to 'Put the spoon on the bed' ('Legg skjeen oppå sengen'). This requires the child to understand locative prepositions and two nouns in relation to each other. Cross-linguistic findings by Fidler and Rus (2007) support the fact that locatives are acquired early, and Mitrofanova and Westergaard (2018) have found the same for acquisition of Norwegian locatives. Thus, this section is placed relatively early in the NRDLs. Furthermore, grammatical complexity seems to have relatively little effect on the acquisition of early spatial language comprehension (Fidler & Rus, 2007).

Gillam, Marquardt and Martin (2001) have identified the important prepositions that are common in children's early two-word utterances. These are high-frequency prepositions such as 'in' and 'on', which are also prepositions that are tested in section (Bii) of the comprehension part. The goal here is for the children to produce the correct preposition together with the noun phrase 'the truck' ('lastebilen').

3.2.4 Verbs

Sentence building is part of both the comprehension and the production parts, and there is much theory that explains the complex processes that constitute the process of learning sentence building. According to Tomasello (1992, 2003) and McClure, Pine and Lieven (2006), children acquire island constructions, such as subject-verb constructions independently of verb-object constructions. The verbs children use in SV constructions versus in VO depends on what they have been exposed to, and might use some verbs in the SV construction only, while others exclusively in the VO construction. The NRDLs tests for both of these constructions together. Moreover, it seems that complete understanding of complete SVO structures is acquired later, and therefore this structure is only found in section

(Dii) in both parts of the NRDLs, and partially section (F) of the production part. According to Akhtar (1999), English-speaking children are fully aware of the SVO order when they are four years old. Norwegian is primarily a V2 language, meaning that in instances where an Adverbial occurs at the beginning of the sentence, the Verb will come directly after, making the structure VSO rather than SVO (Larsen, 2014). Thus, it is debatable whether Norwegian truly can be classified as an SVO language. Nevertheless, in the Norwegian NRDLs all the same items testing the SVO order has been kept, as the particular items in section (Dii) do function in the same way in Norwegian and English. According to Larsen (2014) children of 2:00-2:06 years are able to correctly produce some correct sentences with the V2 structure, and a possible hypothesis might be that when V2 is acquired or is in the process of being acquired, the SVO structure might already be in place, thus possibly explaining the placement of the section testing SVO in the Norwegian NRDLs as well.

In section (E), the NRDLs tests verb inflections. It seems that a delay in the ability to conjugate verbs is a common indicator of a specific language impairment (Rice & Wexler, 2001). Furthermore, Rice and Wexler (2001) note that the ability to conjugate verbs in general is a good indicator for detecting possible language impairments ((S)LLs), as this seems to be a feature that is not affected by the socio-economic status of the parents; thus, it appears to be a good clinical marker (Rice & Wexler, 2001).

An issue with detecting difficulties with verb endings concerns the level of difficulty of recording verb endings in speech. For instance, in English it can be common to use the present while speaking about events in the past and thus it may be difficult to discern whether an individual truly master the past tense inflection or not (Edwards et al., 2011). Furthermore, an individual may know how to use the past tense but may not be able to produce it correctly in a test situation. Nevertheless, the NRDLs may still be a sufficient test instrument to use as a starting point for discovering any potential difficulties with this. According to Valian (2006), at around the age of 2:00 years children have difficulties with suffixes that indicate tense. However, 3:00 year olds seem to have better understanding of this feature.

Nevertheless, the concrete results vary between studies and tests, and presently there is no definite answer as to the exact age children acquire an understanding of verb tense inflections (Valian, 2006).

3.2.5 Pronouns

The comprehension part of the NRDLs has a section on pronouns, section (F). Difficulties with pronouns may be linked to developmental language impairments. Furthermore, according to Van der Lely and Stollwerck (1997) small children will in many languages have difficulties with referring pronouns. This also applies to individuals with aphasia: those who have difficulties with comprehension or speech production due to a stroke, various head injuries or other causes (Edwards & Varlokosta, 2007).

Furthermore, the NRDLs tests for both reflexive and non-reflexive pronouns. There seems to be an asymmetrical development of knowledge of reflexive and non-reflexive pronouns in children (Chien & Wexler, 1990), and children seem to find reflexive pronouns, such as ‘He cleans himself’, with ‘himself’ being a reflexive pronoun, easier than non-reflexive pronouns. In the work with the standardisation of the Norwegian NRDLs, many of the children, especially the younger found items such as item number 49 ‘Is the father washing him?’ (‘Vasker pappa ham?’) difficult compared to items such as item 47 ‘Is the father covering himself?’ (‘Brer pappa teppet over seg selv?’), as more children gave erroneous responses to item 49 and similar items than to item 47 and similar items. According to Van der Lely and Stollwerck (1997) and Guasti (2002), children with (S)LIs may have problems with both reflexive and non-reflexive pronouns, and these problems may last up until the age of 12 years (Van der Lely & Stollwerck 1997; Montgomery & Evans, 2009). However, for children with English as their second language, there seems to be no substantial difficulties acquiring either type of pronouns compared to children with English as their first language (Marinis & Chondrogianni, 2011).

3.2.6 Complex sentence structures

In section (G) in the comprehension part and section (F) in the production part, children are tested for complex sentence constructions such as the passive construction, relative sentences and production of questions. According to Dick, Wulfeck, Krupa-Kwiatkowski and Bates (2004), difficulties with these types of constructions might be an indication of (S)LIs. Guasti (2002) notes that especially the passive construction is challenging for children, and that this is often mastered first when they are around the age of 6 or 7 years. Especially if the passive construction in question does not denote a concrete action, this will be especially challenging for children to understand. Warren (2013) also notes that passives seem to be more difficult to parse, as in some linguistic theories of grammar passivisation is considered a transformation of an underlying active sentence (Warren, 2013).

Furthermore, in section (G) of the comprehension part and section (Fii) of the production part, the NRDLS tests relative sentences. These prove especially difficult for most children in the age range for which the NRDLS is norm-referenced, especially if the relative clause modifies the direct object of the sentence. For instance, a sentence such as ‘I have a friend who is nice’ where the relative clause ‘who is nice’ modifies the object ‘a friend’, could be perceived as more difficult than a sentence such as ‘The man who is tall kicked the ball’, where the relative clause modifies the subject (Edwards et al., 2011). For this particular construction where the relative clause modifies the object, it seems that children do not master it before around the age of 8 years (Guasti 2002). Thus, the NRDLS does not actually test this specific construction, only relative sentences where the clause modifies the subject of the sentence.

In the production part in section (Fi) the child is prompted to elicit WH-questions, such as item 49 ‘Which elephant is carried by the boy?’ (‘Hvilken elefant blir båret av gutten?’) where the correct response requires the object to be fronted. According to Guasti (2002), difficulties with this type of item in this section might have to do with the complexity of the structure itself, and not necessarily any difficulties with WH-questions.

3.2.7 Inferences

The last section of the comprehension part, section (H), has to do with inference. This section is largely unchanged from the previous version of the language scales. For the items in this section, the children look at one large picture with several different people performing various tasks and are then asked questions by the test administrator pertaining to the picture. The children must then use their common knowledge of the world and what they see in the picture in order to infer the answers. Some children may have difficulties with this, especially the younger children. Additionally, children with a pragmatic language impairment may struggle with this section, and individuals with other types of (S)LIs may also experience difficulties (Edwards et al., 2011).

3.2.8 Grammaticality judgement

The last section of the production part, section (G), is grammaticality judgement. Here, a series of short sentences are read to the child, and they must state whether the sentence they are hearing is grammatically correct or not, or simply if it is possible to say or not. This section requires meta-linguistic awareness and is something that is generally difficult for children under the age of around 6 years to master (McDonald, 2008). According to

McDonald (2008), this ability generally develops sometime between the age of 6 and 11 years, and only adults seem to completely master this ability. In the earliest version of the language scales, this section specifically tested changes in word order and leaving out articles. Then regular and irregular plural forms were tested, and then progressive verb forms and question words. However, it was found that even the oldest children did not recognise mistakes in regular and irregular past and 3rd person present singular. In the English version of the NRDLs, this section tests word order, progressive endings, conjugation of verbs in 3rd person singular, irregular plurals and irregular past. The Norwegian NRDLs tests word order, articles, conjugation of verbs in the simple present, irregular plural nouns, and irregular past. The Norwegian NRDLs does not test progressive endings, simply because the Norwegian language does not have this category grammaticalized the same way as in English. In order to represent this aspect in Norwegian, a more complex structure is needed, for instance by adding the VP 'holde på med noe' (Eng: 'be doing something').

When administering section (G) on grammaticality judgement, it is difficult for the test administrators to know exactly what children have trouble with, and if they truly know the answer or not. First, the instructions that are given to the child are as follows: 'Monkey is going to say some things but sometimes he says things that sound wrong. If it sounds wrong, say "No", and if it sounds OK to you, say "Yes".' ('Apekatten skal si noen ting, men iblant kan han si ting som høres feil ut. Hvis det høres feil ut skal du si «Nei» og hvis det høres riktig ut skal du si «Ja»'). Here, it is difficult to know whether a child truly responds to the perceived grammaticality of the sentence, or if they are responding to the perceived possibility of the action or the reality that the sentence describes. For instance, in item 61 'The Monkey flew through the air' ('Apekatten fløy gjennom lufta'), children may respond "no" as they might reason that monkeys cannot fly. Furthermore, as this section requires yes/no answers and the section only contains eight items, it is possible for a child to guess throughout the entire section and still obtain an acceptable score. Nevertheless, if there are any concerns regarding a child's performance on this section, further testing using other test instruments may be necessary.

In conclusion, there are a series of features that are associated with developmental language impairments in children (and adults), and the NRDLs tests the acquisition of all of these features, in order to discern whether there may be cause for further testing of certain features an individual might have difficulties with.

4.0 Adapting clinical instruments into different languages

4.1 Generally on test adaptation

When it comes to using an existing test or evaluation instrument in a different language than the language in which it was originally developed, we commonly speak about ‘test adaptation’ rather than ‘test translation’. There seems to be a common notion that translation denotes the direct and often literal transference of words from the source language into a target language (Hambleton, Merenda & Spielberger, 2004). However, when adapting test instruments, it is not sufficient to simply translate the question items and the possible responses: both of these must also be adapted to fit the cultural context of the target culture. Thus, when translating a test instrument into a different language than its original, common practice is to talk about ‘test adaptation’ as it includes a broader process than just translation (Hambleton et al., 2004).

According to Hambleton et al. (2004), poorly adapted tests can in many instances lead to less than ideal, or even invalid research, as suggested by some cross-cultural research. This may be due to a number of reasons, such as bad understanding of the target culture or of what constructs the test measures. There is a number of concerns that must be considered when adapting tests, and Hambleton et al. (2004) point out three main sources of errors that are common in test adaptation:

1. Cultural or linguistic differences
2. Technical issues, design and method
3. Interpretation of results

The first issue relating to cultural or linguistic differences mainly concerns the adaptation in itself, before the administration of the test has begun. When adapting a standardised test into another language, the key idea is to use concepts in the source language and find the equivalent in the target language. However, the notion of equivalence becomes especially questionable when test adaptation is concerned, as Van de Vijver and Poortinga (2004) note. A critical question is whether it is even possible to achieve perfect equivalence between source and target text when performing a test adaptation. Nevertheless, test adaptation requires virtually perfect command of both the source and the target language (Van de Vijver, & Poortinga, 2004; Bachman & Palmer, 1996).

Regarding the second point Hambleton et al. (2004) mention, technical issues, design and method, this might also relate to biases, and especially construct bias, which is arguably the most crucial point to be aware of when adapting clinical language tests such as the NRDLs. As Brislin (1986) notes when adapting tests into different languages, is the importance of the translator being familiar with not only the material that they are adapting, but also the intrapsychology of the tool – what exactly is it developed to measure, what effects are expected, and how can it be adapted into a different language while preserving these aspects? Moreover, test translators should also be familiar with all empirical connotations of the test instrument, and they must also be familiar with all linguistic and literal referents it uses. To sum up, the translator cannot simply be competent in both the source and the target language they are working with; they must also be somewhat familiar with the test instrument and its function, or the very least familiar with the implications of test adaptation (Brislin, 1986). This second point also encompasses the possibility of using back-translation or decentring.

Both Hambleton et al. (2004) and Brislin (1986) note that the preferable method for test adaptation is translating the test into the target language, and then back-translating it into the source language, before comparing the original and the back-translated versions. However, as Hambleton et al. (2004) note, this method might not pick up subtle errors or inadequate translations such as single translated words or items that would not work well in the target culture context, but when it is back-translated it appears to work well. Brislin (1986) also talks about the concept of decentring, which essentially denotes the process of moving the test development away from one culture or language, and instead developing the same test instruments simultaneously in two or more languages where it is intended to be used. While this appears to be one of the most beneficial methods for test adaptation as it would better ensure that there is less cultural bias or construct bias, this also appears to be a relatively difficult feat to manage in practice, both regarding finances and simple practical considerations. Furthermore, test instruments such as the NRDLs and similar clinical tools are usually a product of years of development, data collection and revisions, and developing this tool in collaboration with one or more other research communities in other countries would surely have been slightly impractical. Nevertheless, decentring would presumably be an effective method for smaller-scale research projects.

The third point Hambleton et al. (2004) mention, interpretation of results, have to do with certain biases regarding how the results of a test are interpreted after testing. This is beyond

the scope of this thesis and is not as relevant concerning the adaptation of the NRDLs, as it is generally more applicable for standardised tests used in education. However, this can essentially be affected by differences in curricula, differences in motivations in students who take the test, and other various socio-political factors (Hambleton et al., 2004).

4.1.1 Guidelines for test development

Brislin (1986) provides a list of twelve essential considerations for developing or modifying test items, in order to ensure that they are easily adaptable for later use, and that key components are kept in the adaptations.

The first guideline he mentions is to use short and simple sentences with less than sixteen words in them, in order to ensure that there is no more than one key idea per sentence. However, test items may of course be longer than one sentence. The second guideline is to use the active voice rather than the passive as far as it is possible, as it tends to be easier to translate active sentences, and how the passive voice is realised differs greatly between languages (Siewierska, 2013). The third guideline is to repeat nouns rather than replacing them by pronouns. This is in order to make it clearer who the sentence is referring to in a text. The fourth guideline is to avoid metaphors or colloquialisms, as these are usually not universal, might lose their meaning in adaptations, and may be a point of confusion to the reader if they are kept in a language where the original meaning behind them has not been adequately transferred.

The fifth guideline is to avoid the subjunctive and modal verbs such as could, would and should. These also prove difficult to translate and might even have no clear equivalent in the target language. The sixth guideline is to add sentences to provide context and elaborate on key ideas if necessary. The seventh guideline is to avoid prepositions and adverbials that indicate time and place, as these concepts might not always be universal and might not have a direct equivalent in every language. Examples are adverbs such as 'frequently', 'often', and 'seldom'. There might be cultural, as well as individual differences in the understanding of these terms that are not especially specific. The eighth guideline is to avoid using possessives, as there might be different concepts of ownership across cultures.

The ninth guideline is to try to be as specific as possible. Using specific rather than general terms might prove useful, as people are specific and are familiar with specific concepts and specific terms in their own cultures. As Brislin (1986) notes, it is beneficial to use decentring

and find concepts that might be familiar to all groups of people, or concepts that can be specified to differently in different groups, but still remain relatively similar across groups. The tenth guideline is to be as clear as possible and avoid vagueness. More specifically this means avoiding using words such as ‘maybe’ and ‘probably’, as these bear no concrete meaning and may have different connotations or meanings across cultures. The eleventh guideline for test development is to use wording that can be expected to be familiar to translators. If possible, the researcher or clinicians developing the test instrument should be working together with a translator to develop the textual materials. If the translator points out that a word that is suggested has no actual equivalent in the target language, it would be wise to use another word, or alter the item altogether in order to avoid the specific word. The twelfth and last item is to avoid using two different verbs in a single sentence that can suggest two different actions. If it appears that one or more test takers are struggling with this item, it will be difficult to know what exact part of the sentence test takers are having trouble with.

To sum up, what all of these guidelines have in common is that they stress the use of clear and concise language and emphasise that using language or examples that are too specific to one culture should be avoided as much as possible, in order to ease the process of future adaptation into different cultural contexts.

4.1.2 Guidelines for test adaptation

In addition to Brislin’s (1986) guidelines for test development, a series of 22 guidelines for test adaptation are presented in Hambleton (1994) and Van de Vijver and Hambleton (1996).

The first two guidelines are related to the context of the test. The first guideline is to strive to minimise the effects of cultural differences where this is not relevant for the overall test. In most cases it will not be possible or desirable to completely eradicate all the effects of cultural differences, therefore they have worded it as ‘minimise’. The second guideline is to assess construct equivalence. Construct equivalence essentially denotes the notion that a construct is supposed to measure the same construct in the two different language cultures. If construct equivalence is achieved, ideally, similar data should be produced by the two versions of a test, even though the items are not necessarily similar in their wording.

The ten next guidelines concern the test development and adaptation. The third guideline is to ensure that full account has been taken of any cultural and linguistic differences between the source and target language. An example Hambleton (1994) uses is the translation of ‘webbed

feet' from English to Swedish. The question item was to identify the animal that has 'webbed feet', but as the Swedish test had this phrase translated to the equivalent of 'swimming feet', this item became substantially less difficult in Swedish than in English.

The fourth guideline is to provide concrete evidence that the language that is used in all parts of the test is appropriate in the cultural context, as referred to in Brislin (1986). See section 4.1.1 for a detailed explanation of Brislin's (1986) guidelines for wording of research instruments and tests. The fifth guideline is to provide evidence that the format that is used is appropriate for and familiar to the intended group of test takers. Similarly, the sixth guideline is that evidence should be provided that the stimulus and item contents are familiar to the intended audience. The seventh guideline is to compile judgemental evidence of the validity of the adapted text, for instance by providing a back translation, or use more than one translator. The eighth guideline is to ensure that the data collection design allows for the use of suitable statistical techniques to establish construct and item equivalence between tests. The ninth guideline is to apply these statistical techniques to establish the equivalence between language tests and identify components that may potentially pose problems. The tenth guideline is to provide information on the validity of the adapted test in the population for which the test is intended. Even though the original test has high validity, this might not be the case for the adapted version, due to for instance poor adaptation, or simply the fact that construct equivalence is not possible to achieve as there is insufficient construct validity for the construct in the target language culture. The eleventh guideline is to provide statistical evidence for item equivalence in all the populations for which the test is intended. In brief terms, item equivalence essentially means that an item garners the same data from the respective populations tested on the item in both the original and the adapted version (Sireci & Allalouf, 2003). The twelfth guideline states that although items that prove to be non-equivalent cannot be used directly to compare or link two versions of a test, these items may be used to provide information about one specific population. Thus, items that are non-equivalent in two versions do not necessarily need to be automatically discarded.

The next six guidelines relate to the implementation and administration of the test. The thirteenth guideline is that all conditions that may affect performance should be as similar as possible across cultures. The fourteenth guideline is to try to predict the types of issues that might occur and attempt to account for these. An example of this is how to give test instructions. The fifteenth guideline is that the test administrators must be sensitive to all

issues relating to stimulus materials and administration procedures. In other words, the test administrator must be familiar with all test materials and items, must know the procedures and possible issues that can occur, and the purpose and procedures of the testing. The sixteenth guideline is that the instructions for test administration should be available in both the source and the target language. The seventeenth guideline is that the test manual should address all issues that may be scrutinised later when adapting a test. The eighteenth guideline is that the test administrator should be as unobtrusive as possible. The administrator should be aware that factors such as age, gender, cultural background and other background conditions may affect the outcome of the testing and may lead to a so-called interviewer effect (Yang & Yu, 2008). Especially if there are considerable apparent differences between the test administrator and the individual being tested in the aforementioned categories, the interviewer effect may be especially apparent.

The last four guidelines relate to the test documentation and interpretation of test scores. The nineteenth guideline is that when a test is adapted, documentation on change and evidence that supports the equivalence between the source and the target text as a basis for these changes should always be provided. The twentieth guideline is that differences between samples from different populations should not be taken at face value, and other factors should also be controlled. The twenty-first guideline is that it is only possible to compare results across populations at the level of invariance that has already been established for the scale that is used to report the scores. Essentially, the main point is to not make groundless comparisons of scores across languages. If bias is detected, then individual differences within one or several groups cannot be measured and compared with individual differences across groups. However, if there is no bias, it is possible for individual differences within and across groups to be measured at the same level. The twenty-second guideline is that the test developers should provide information on certain factors such as socio-political aspects that may affect the results and provide possible solutions to account for these factors.

In sum, all of these guidelines stress the importance of familiarity with both the source and target language as well as the implications of the test instrument, and the importance of clarity, validity and providing proof of all effects the test produces.

4.2 Developing clinical tools into different languages

Sireci and Allalouf (2003) point out that there is growing need for adaptation of existing test instruments, as for instance standardised tests in secondary education in the United States see the need of test instructions in different languages so that the tests may be administered to a growing population of people not having English as their L1. In general, in a society that is becoming more multicultural and thereby also more multilingual there is a growing need of writing not only standardised tests, but also texts such as educational material, public notices and other texts in general, in several languages in order make public services, education and other central functions in society more accessible to a greater number of people. Additionally, clinical tests can be adapted into other languages, as they usually are tools that have been standardised and well-tested before publication, and might therefore be useful in other languages as well. Two examples of clinical test tool that have been adapted into different languages, in addition to the NRDLs, is TROG-2 and the ASQ.

4.2.1 TROG-2

Test for Reception of Grammar, or TROG-2, is a test for receptive language, and it is intended for clinical and scientific use, specifically by speech therapists, psychologists and language teachers working with people who are deaf or hearing-impaired. It is intended to be appropriate for a wide age range, from 4:00 years to adults. More specifically, it can be used on individuals with suspected developmental language impairments, hearing loss, physical handicaps that affect speech production, learning difficulties, or acquired aphasia (Bishop, 2003). As a further note, similar to the NRDLs, TROG-2 is not intended to be used as a tool for final diagnostic of language impairments, but can be used as a diagnostic tool to justify further examination. The original TROG was developed in the 1970s, and was not intended for clinical use (Bishop, 2003). The current version, TROG-2, was published in 2003, and was standardised on a large sample of children and adults in the UK and Australia (Bishop, 2003).

The test consists of a total of 80 items, and the items are four-choice, wherein one is correct. The test contains a stimulus book, a manual, and a protocol detailing each item with scoring. As the test is intended for children as young as 4 years and individuals with language impairments, the vocabulary of the test is restricted to a limited set of simple nouns, verbs and adjectives. The test examines several grammatical concepts, and each of these concepts are divided into blocks of four, making up a total of 20 blocks. The blocks are arranged by

difficulty, so that each block gets increasingly difficult. The test should be ended when an individual fails five consecutive blocks in a row (Bishop, 2003).

Additionally, TROG-2 has an optional section for basic vocabulary items. This section might be administered before or after the main section as the test administrator sees fit. This section is divided into eight pages with six images on each page. The images represent basic vocabulary that is used in the remainder of the test, and this section can be administered to ensure that the individual being tested understands the vocabulary the test uses. There are several ways of administering this section, but the easiest way would be to have the test administrator name the images on the page and then have the individual that is being tested point to the relevant image. It is also possible to have them name each picture, but contrary to the NRDLs the TROG-2 does not actually require productive language, so this might not always be a viable option.

TROG-2 for use with English-language individuals has been standardised on a total of 792 children in the age range between 4:00 and 16:00 years, and on a total of 70 adults. For the adaptation of the Norwegian version of TROG-2, the test and the manual were published in 2009. It has been standardised on a total of 950 Norwegian children and teenagers in the age group between 4:00 and 16:04. There is both a Bokmål and Nynorsk version of the test, although there are only marginal differences between these two versions regarding content and test items. According to tests for reliability, the Norwegian TROG-2 proves to have high inner consistency, with a Cronbach's alpha of 0.95 (Bishop, 2009).

Lyster (2008) also mentions that in general, the standardisation of the Norwegian TROG-2 proved to provide similar results as the original British version, and that British and Norwegian children generally have the same grammatical competence. One difference was that Norwegian children performed slightly better in the age bracket between 4 and 6 years (Lyster, 2008). Lyster (2008) suggests that this might be due to the fact that English-speaking children must acquire two different present tenses, both simple present and present continuous, whereas Norwegian children only use the simple present.

4.2.2 ASQ

The Ages and Stages Questionnaire (ASQ) is a questionnaire for parents to fill out regarding the development of their child, especially with regards to their emotional and social competence, motor skills and problem-solving skills (Tsai et al., 2006). The questionnaire is

intended for children between the ages of 0:04 and 5:00 years. There is a total of 19 questionnaires with 30 questions each, and the questionnaires vary depending on the age of the child.

In a study from 2006, Tsai et al. looked at the adaptation of the ASQ in Taiwan, with regards to cultural appropriateness, inter-rater reliability, internal consistency, criterion validity and content validity. For the test adaptation, the 36-month version of the ASQ was translated from English into Mandarin Chinese by a native speaker, then back-translated into English by an English-Mandarin interpreter. When the two translations were compared, no apparent discrepancies were found, indicating that the translation of the text was adequate (Tsai et al., 2006). Additionally, an expert panel of five people reviewed the adapted version with regards to cultural appropriateness for the Taiwanese cultural context. The questionnaire was also distributed to 35 different pre-schools in southern Taiwan, who had parents and teachers of a total of 112 children fill out the adapted questionnaire. The main findings of this preliminary study were that the adaptation appeared to be culturally appropriate for the Taiwanese context, and there also appeared to be good internal consistency. The adapted questionnaire also managed to identify all the children that had previously been identified as having developmental delays (Tsai et al., 2006).

As both Tsai et al. (2006) and Van Heerden, Hsiao, Matafwali, Louw, and Richter (2017) have found, the ASQ is an important low-cost tool that is easy to use for parents, as it does not require any substantial training, and only approximately a 6th grade reading level. The ASQ can be an important diagnostic tool for identifying young children who are in need of early intervention, especially in low- and middle-income countries.

Tsai et al. (2006) estimated that at the time of writing, around 80% of children with developmental delays were not identified in Taiwan. Some causes might be both lacking health services and tools for correct identification of such delays, as well as a certain societal stigma associated with developmental delays in children. Nevertheless, the ASQ is an effective instrument with the possible ability to compensate for the previous unsuccessful identification of children with developmental delays.

The ASQ has also been adapted and standardised for use in Norway. There are already existing, functioning screening processes that are open for all families with babies in Norway

that can identify developmental delays in children, but the ASQ may still function as a complement and supplement to the existing services (Janson & Squires, 2004).

5.0 The Norwegian NRDLs

5.1 The standardisation process

In the original plan for the adaptation and standardisation of the Norwegian NRDLs, the preliminary plan was to adapt the test in January-February 2017, complete a pilot study in 2017, and then complete the standardisation and data collection by the end of the spring of 2018. The goal was to test a total of 660 children, equally distributed across sex and age intervals.

In reality, the piloting of the test took place between November 2017 and March 2018. A total of 50 children in the age span between 2:00 and 7:00 years, 25 boys and 25 girls, were tested. The standardisation and collection of normative data for the NRDLs took place between April 2018 and November 2018. A total of 242 children were tested in each 6-month age bracket between 3:00 and 7:10 years, 120 boys and 122 girls.

Before participating, all the parents were given complete information on what the collected data was intended for, so that they were able to provide informed consent on behalf of their children (see appendix A and B). They could withdraw from the experiment at any point. Furthermore, the children could also choose not to participate if they wished to, which a few individuals did. The project was formally approved by the municipality of Trondheim, and NSD (The Norwegian Centre for Research Data).

Through the process of standardisation and collection of normative data, the test administrators and adaptors encountered several challenges with regards to recruitment of informants, scoring of one of the items, and participation of the youngest children, which will be discussed in detail in the next sub-sections. Additionally, it was found that there are slight gender differences in Norwegian children's performance on the NRDLs, which will be discussed in section 5.2.

5.1.1 Recruiting informants

When working on the standardisation and collection of data for the Norwegian version of the NRDLs, the first part was piloting the test which lasted from November 2017 to March 2018. A total of 50 children between the ages of 2:00 and 7:00 took part in the pilot-testing, 25 males and 25 females. The informants were recruited through schools and kindergartens in the Trondheim area. Through the piloting, some of the wording of the test items was revised and changed, and the test administrators practiced the administration of the NRDLs to

children of different ages. For the piloting, six different test administrators were trained and conducted the data collection.

The process of collecting data for the norm-referencing of the NRDLS began in the spring of 2018 and was completed in November 2018. A total of 242 children took part in the testing, 120 males and 122 females. The ages of the participants ranged from 3:00-7:10, distributed across 9 age brackets (see appendix D, table 1.0). The participants were recruited from several schools and kindergartens in central Eastern Norway, and the wider Trøndelag area. Participants were excluded if they had any known language impairments, impaired hearing, other types of functional impairments, or did not have Norwegian as one of their native languages. For the standardisation, four of the six test administrators collected the data.

The process of recruitment of participants proved to be relatively difficult. The nature of the data collection required the team to first come in contact with relevant schools and kindergartens. Then the schools and kindergartens had to agree to participate in the project, before they would distribute the consent forms to the parents of the children who could participate. The problems arose when several of the institutions that were approached declined to participate altogether, as many expressed that it would take too much time and that they did not have sufficient resources. An interesting contrast to a similar project as presented by Tsai et al. (2006), where the ASQ was developed for use in Taiwan. The authors describe that out of the 35 preschools they contacted, 35 preschools agreed to participate. For the Norwegian team, several schools and kindergartens were contacted, but only some agreed to participate as many declined on the basis of time constraints, resources, or simply refrained from replying. There is perhaps something to be said about cultural differences and the different attitudes and the feeling of responsibility towards authority figures and institutions of research (Jacobsen & Bruun, 2000), although that is beyond the scope of this thesis. As Trondheim has The Norwegian University of Science and Technology as a prominent institution, a large number of the schools and kindergartens in the Trondheim area were already taking part in several other research projects, something that was also expressed by a number of the institutions the Norwegian team approached when they were stating the reasons for why they did not wish to participate.

Another issue was the simple fact that the consent had to go through the parents or the guardians of the children, and this created an additional step wherein it is a possibility that many parents forgot to return the consent form, and thus it was not possible to test their child.

In general, recruiting informants may prove to be difficult, as there naturally are measures in place to protect the privacy and personal information of individuals, and especially children, and which requires participants to give informed consent. Furthermore, as parents also need to be consulted, there is the additional step that must be taken before the data collection can commence.

5.1.2 Item number 25 and object permanence

In section (Ci) of the comprehension part of the NRDLS, in item number 25 the test administrator instructs the child to ‘make Monkey hide’ (‘få apekatten til å gå og gjemme seg’). In the manual, the instructions explicitly say that the child will only score a point if they make Monkey go hide somewhere that is out of sight of the test administrator (under the table, behind the child etc). It also explicitly says that if the child makes Monkey hide by putting his hands in front of his eyes, this will result in zero points for that test item (Edwards et al., 2011).

However, this seems to open up an interesting question regarding both children’s development of object permanence and also their understanding of other people’s perceptions. Object permanence describes the notion that objects continue to exist even when they are out of perception (Baillargeon & DeVos, 1991). As observed by the test administrators, a good portion of the younger children, particularly in the 3:00-3:05 age range, would actually make the monkey go hide by putting his hands in front of his eyes. It could be argued that the children choosing this approach seemingly may have not yet properly developed object permanence. Object permanence, as described by Piaget, is generally believed to develop around the end of the sensorimotor stage, so around the age of two to three years (Piaget, Vonèche, & Gruber, 1995). However, as described by for instance Baillargeon and DeVos (1991), infants as young as three months old may have some sense of object permanence. Thus, it would be difficult to argue that the children aged 3:00 years and up choose to solve the task in this manner because they have no sense of object permanence.

Another explanation for why some children might choose this approach for the monkey might be that the children perceive the monkey to be a character that is both somewhat childlike or silly, and also slightly playful. This might be in line with what Aguiar and Taylor (2015) have found, namely that children tend to view stuffed animals as having some agency. Some children might therefore make him try to hide in a manner they know is not sufficient, but that they perceive the monkey to believe works. Furthermore, this image of the monkey

as a somewhat ‘silly character’ is reinforced in later sections of the NRDLs, for instance in section (F) of the comprehension part, when the test administrator tells the child that ‘Now Monkey is going to ask you some questions about some pictures. Sometimes he might try to trick you’ (‘Nå skal apekatten spørre deg litt om disse bildene. Det kan hende han prøver å tulle med deg og sier feil’). This might possibly reinforce the image of Monkey as a somewhat playful or childish character, who nevertheless has agency.

Lastly, it could be argued that children selecting this approach do still seem to understand the meaning of the verb ‘hide’, which is the main focus of the NRDLs. Furthermore, many children may be accustomed to have their parents play a similar game with them or younger siblings where the parents ‘hide’ by putting their hands over their eyes (peekaboo).

Nevertheless, the reasoning behind the scoring of this item might be that the children that choose this approach have not yet developed the complete understanding of the semantics of the verb ‘hide’, which essentially would be something along the lines of ‘to put out of sight’, or ‘be out of sight’ (hide, *Merriam-Webster*, 2019). However, the main objective of the NRDLs is to measure language comprehension and production, and although the child does not make the monkey hide in the sense that he is out of sight, they still perform an action that might indicate that they have partially understood what they believe to be the meaning of ‘hide’, and they have demonstrated one possible instance where this verb can be used.

5.1.3 Difficulties with children under 3:00 years

In the original English language manual for the NRDLs, the authors recommend that the instrument can be used for children between the ages of 1:06 and 7:00 years of age. In the process of collecting the data for the norm-reference for the Swedish version of the test, children between the ages of 2:00 and 6:11 participated. Similarly, for the development of the Norwegian version of the NRDLs, children between the ages of 2:00 and 7:05 were recruited. However, as the data collection proceeded, it became apparent that it was nearly impossible to make the children under the age of 3:00 willingly participate in the test, not to mention to elicit any useful responses from them. As the test administrators observed, they would either refuse to point to the relevant objects when prompted by the test administrators or fail to name the objects presented by the test administrators. As all of the children aged 2:00-2:11 were accompanied by either a parent or an adult that worked in their kindergarten, the test administrators would usually ask the parents or caretakers if this was normal behaviour for the child in question. They would then respond something along the lines of

‘usually, she knows this word’ or ‘he can actually talk, but it does not seem like he wants to right now’.

The question then arises regarding why the Norwegian children under the age of 3:00 either were completely unwilling to participate or failed to indicate that they understood the relevant tasks. From the Swedish (Hammarström, Kjellmer & Hansson, 2017) and English manuals (Edwards et al., 2011) it appears that both the Swedish and the English test administrators managed to elicit responses from the youngest children aged 2:00-2:11 years .

A possible explanation for why the Norwegian children did not seem to cooperate is that they are not familiar with the type of formal testing situation that the NRDLs presents, or they are not familiar with producing responses in this type of context. However, as all of the children attended kindergarten on a regular basis, it seems inaccurate to assume that they were not familiar with a situation where they are asked to point out objects or images or respond to questions such as ‘Where are your ears?’ (‘Hvor er ørene dine?’). Furthermore, as most of the parents or caretakers with the children would stress that their non-responsiveness was an example of abnormal behaviour for the child in question, it might be relevant to look at the general developmental patterns of children at this age. According to Berk (2012), children generally develop a sense of agency and a stronger desire to be independent between the ages of 2:00 and 3:00, and thus might be more likely to resist cooperating in situations they simply do not feel any desire to partake in.

Another possible factor that might have affected the result is the sample size and the composition of the population. In the development of the English version of the NRDLs, results from a total of 1266 children were used, while in the development of the Swedish version, results from 530 children were used. However, due to financial and time constraints, data from no more than 242 children for the Norwegian version was available. Furthermore, although the NRDLs manual provides explicit guidelines for when to end a session and how children should be encouraged to provide a reply, there might also have been differences in how the test was administered in the different countries. Moreover, there is naturally also the factor of the mood and the level of energy in each individual child that might explain some of the lack of cooperation.

Lastly, another possibility is that the presence of a parent or another caregiver was part of the reason why the children in this age group were highly uncooperative. All of the other children

who participated without an additional adult were for the most part highly cooperative, and performed more or less as expected, even the ones in the age group 3:00-3:05. However, it is questionable whether it would be beneficial to test children as young as 2:00-2:05 without an adult, as the comfort and the feeling of security for the child is a crucial concern for the administration of the test, and for a very young child to be left alone in a room with an unfamiliar adult might not be a reasonable solution.

For further work with the NRDLs in Norway, it might be of interest to attempt to take a further look at children aged 2:00-2:11, look at for instance more dialect variants or even children with other L1s than Norwegian, and examine whether it truly is impossible to elicit useful responses from this demographic of Norwegian children, or if there are any other variables that might influence the results. There is no doubt that typically developing children between the ages of 2:00 and 3:00 years usually will talk, as Larsen (2014) among others clearly demonstrates. However, the question remains how to elicit useful responses from these children.

5.2 Gender differences

There were little significant differences between the genders' performance on the standardisation of the Norwegian NRDLs. However, for the comprehension part there was significant difference in the 3:00-3:11 age bracket with a p-value of 0.02, and in the 4:00-4:05 age bracket, with a p-value of 0.04. For the production part, there was a significant difference in the 3:06-3:11 age bracket with a p-value of <0.01, and in the 7:00-7:10 age bracket with a p-value of 0.02. In each of these age brackets the girls performed slightly better than the boys.

According to Edwards et al. (2011), there was also a slight gender difference in the performances of the children during the standardisation of the English NRDLs, the girls performing slightly better than the boys, with a p-value of <0.5. For the Swedish standardisation of the NRDLs, the authors also report a slight advantage for the girls. In the comprehension part there was a significant difference in the age bracket 3:06-3:11 with a p-value of 0.022, and the 5:00-5:05 age bracket with a p-value of 0.006. For the production part, there was a significant difference in the 5:00-5:05 age bracket with a p-value of 0.012, and in the 5:06-5:11 age bracket with a p-value of 0.030 (Hammarström et al., 2017).

Although the gender differences for all three versions of the NRDLS are small, there seems to be a persistent advantage with the girls' performance. Two studies by Eriksson et al. (2012) and Dabašinskienė (2012) seem to reinforce this finding.

5.2.1 Gender differences in ten different language communities

According to a study by Eriksson et al. (2012) where they compared results from studies in ten different language communities, it seems that in general, girls do perform slightly better than boys in some linguistic tasks. In general, girls perform better in tasks concerning gestures, vocabulary comprehension and vocabulary production. The age of the children in this study was 0:08-2:06, making them younger than most of the children in the NRDLS.

There seems to be several theories for why girls seem to perform slightly above boys in language development, and most are attributed to social explanations. Caldera, Houston and O'Brian (1989) argue that the difference is emphasised by the different types of toys boys and girls typically are given to play with. Typically, boys play with toys that are more action-oriented, such as cars, trucks etc., and these do not stimulate language to the same extent that toys intended primarily for girls do. Girls are typically given more care-oriented toys to play with, such as dolls. Furthermore, Bornstein, Haynes, Pascual, Painter and Galperin (1999) claim that in general, parents of girls engage in much more so-called symbolic play than parents of boys, and it seems that symbolic play in children from a young age is strongly linked to advanced language (Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Lytinen, Poikkeus & Laakso, 1997). There is also the argument for more variability in cognitive functioning in men, or the greater male variability hypothesis, or that in the male population there is greater variability in intelligence. There is some variable evidence for this hypothesis, and no definite conclusion has been reached (Eriksson et al., 2012).

In the study presented by Eriksson and colleagues (2012), they note that neurodevelopmental factors such as earlier brain lateralisation in girls, as described by for instance Crow (1998), is a factor that does not actually differ across language communities, but rather is consistent independent of language community, and thus the finding that girls perform at a slightly higher level than boys cannot be attributed to differing brain lateralisation across language communities.

The MacArthur-Bates CDI questionnaire was used to collect the data from a total of 13783 children from 10 different European language communities. The children were between the ages of 0:08 and 2:06.

The main findings overall for the infants were that girls and older infants were able to use and comprehend a larger number of different communicative items than the boys and the younger infants. Furthermore, the language skills of the infants were found to vary between language communities. For communicative gestures, they found that girls seemed to use more different types of communicative gestures than boys. In general, there was no significant interaction between gender of the children and language community, meaning that the differences in use and comprehension of communicative gestures that were found between the genders was consistent across language communities.

For word comprehension, there was no statistically significant effect of gender, and there was no interaction between gender and language community or gender and age group. Thus, for word comprehension, there seems to be no significant difference in performance across genders. For word production, Eriksson et al. (2012) found that girls were actually more variable than boys, giving the greater male variability hypothesis no support.

In general, the infants there was a significant difference between boys and girls in communicative gestures and word production, where the girls had a slight advantage over the boys in all but one case; the Austrian boys produced slightly more words than Austrian girls.

Regarding the toddlers and word production, they found that girls tend to use more different types of words than boys. Interestingly, the difference between boys and girls also seemed to increase with age. This also supports what was found in the collection of the data for the Norwegian NRDLs, that the slight advantage of girls in both word production and word comprehension seem to persist to some degree up until at least 4:05 years, and possibly beyond this point, at least for production. In general, Eriksson et al. (2012) found that boys dominated the lower tail of the ANOVA analysis, whereas the girls prevailed in the upper tail.

It seems that the emerging differences in language skills between the genders where girls tend to dominate, is an effect which seems to occur across language communities, so that this difference is more universal and not dependent on specific cultural or linguistic factors present in certain communities.

No significant differences were found between the genders regarding word comprehension, but the authors note that the MacArthur-Bates CDI might be an inaccurate instrument for determining if toddlers and infants actually have understood and acquired a word or not, as it relies on parental reports. It is possible to make a case that the NRDLs would be a more suitable tool for measuring language comprehension in children, as it takes a more interactive approach. However, the NRDLs is not intended for children under 2:00 years, and typically it might be somewhat difficult to use this instrument on young toddlers and infants.

Nevertheless, this might be the case that the use of an inaccurate measurement tool leads to erroneous results, and that it is in fact the case that girls also perform slightly better than boys in word comprehension. This effect was indeed observed in some of the data from the children in the Norwegian NRDLs.

Lastly, it should be noted that although the differences between boys and girls in language production and communicative gestures seem to be significant, the differences still remain small. The fact that the MacArthur-Bates CDI does not use data that is collected directly from the children but is rather a questionnaire that parents fill out, may be a case for why it might be useful to conduct additional testing in order to have confirm or refute the present effect. On the other hand, the NRDLs use data that is based on observations of direct performance by children, and as this data seem to indicate that certain gender differences are present, this could be taken as a confirmation of the claims posed by Eriksson et al. (2012). The authors conclude the paper by asking if separate norms should be developed for each gender, in order to better reflect the performance of a child of one gender compared to the norm for both the age group and gender. There are some valid arguments for this practice, such as the fact that gender neutral norms could result in more false negatives for girls and false positives for boys. Eriksson et al. (2012) note that this would be especially disadvantageous for girls as it is less ideal to be in need of help and not receiving it. As for the development of the Norwegian norms for the NRDLs, separate norms for each gender would be a possibility, but for the present the resources, time, and number of participants do not allow for this. This is possibly something that could be done in future developments of the NRDLs, both in Norway and other language communities, as it might be beneficial for the children being screened, as well as parents, clinicians, and researchers.

5.2.2 Case study on gender and diminutives

Another case study by Dabašinskienė (2012) looked at the gender differences in acquisition of a more specific linguistic concept, namely Lithuanian diminutives. Generally, diminutives and

hypocoristics in those languages that these concepts play an important role, for instance in Slavic languages such as Lithuanian and Russian, tend to play an important role in early language acquisition (Dabašinskienė, 2012). Diminutives are essentially terms for expressing endearment and affection, or smallness. Hypocoristics refer to this concept with names, essentially the use of pet names or shortened forms of names (Dabašinskienė, 2012). Diminutives seem to be one of the first derivations that children acquire at around 2 to 3 years, making them a crucial indicator of language acquisition and language competence in early child language development (Dabašinskienė, 2012). Diminutives are also generally important for the development of the grammar of a child's language. In general, there seems to be a trend that girls use more diminutives than boys do (at least for data in Russian, see Protassova & Voeikova, 2007), and that parents of girls use more diminutives in the input for the children than parents of boys do (Dabašinskienė, 2012).

In the present study, the researcher followed a Lithuanian girl and a Lithuanian boy for a period of 12 months, during the age intervals between 1;8-2;7 and 1;6-2;6 respectively. The main findings were that the girl used a significantly higher number of diminutives than the boy did: for the girl, the average number of diminutives in her speech was between 32 and 48%, whereas in the boy's language the average amount of diminutives was between 15 and 25%, slightly lower.

Dabašinskienė (2012) explains this primarily by pointing to the number of diminutives in the input from the parents. Similarly, the parents of the boy use significantly less diminutives than the parents of the girl: between 18 and 35% for the mother of the boy, in contrast with between 25 and 51% for the mother of the girl. Furthermore, Dabašinskienė (2012) found that the girl in general used the diminutive suffixes more creatively and spontaneously than the boy did. This finding might also correlate with the number of diminutives in the input from the parents.

In general, most likely it is not possible to say that this study can be taken to generalise a larger population, as only two children were studied in this case study. Nevertheless, the results can tell us something about the fact that input seem to strongly correlate with output in young children. Furthermore, it might say something about the fact that the amount and quality of the input affects how language is produced, and that there seems to be significant differences in how parents address and talk with their children depending on the gender of the child. This may also support the findings from the Norwegian standardisation of the NRDLs and may also be an explanation for the difference in performance depending on gender. However, it is not possible to say definitely if any differences in performance is due to societal constructs and differences in how parents raise their children, how they speak to them and what types of toys they are encouraged to play with, or if

there are any concrete biological or neurodevelopmental underlying causes. Nevertheless, more research is needed in order to determine the true extent of these possible gender differences in language acquisition and language performance. This is also the case for the Norwegian NRDLs, as more research is needed to understand the true extent of the slight gender differences observed from the present data sample.

5.3 The adaptation from Norwegian into English

Norwegian and English are both Germanic languages, and thus share many features (Vulchanova et al., 2019). However, there are certain differences on the word formation level, on the syntactic level, as well as with certain verb structures, such as passive formation (Åfarli, 1989). The present tense is also used differently in Norwegian and English; while in English the present progressive is commonly used when talking about an action happening right now (Hasselgård, Lysvåg & Johansson, 2012), in Norwegian progressive aspect is not grammaticalised in the same manner as in English (see section 3.1), and the simple present is what is normally used for events or actions taking place in the moment (Vulchanova et al., 2019).

For the adaptation of the items and the instructions, the test adaptors strived to stay as close to the original as possible, by keeping all items, objects and images that the original use.

However, as mentioned earlier, since there are slight differences in the Norwegian grammar, some wording was changed in order to better correspond with how Norwegian children speak and how the language input they get on a daily basis is. For instance, all dialect variants of words and verb conjugations are accepted (Vulchanova et al., 2019). As Norway is a country of a diverse linguistic landscape with many different dialects, children will necessarily have been exposed to specific dialect variants and words as they grow up (Røynealand, 2009).

As is the case for the adaptation of the Norwegian NRDLs, the most important aspect was not in fact ensuring that each item corresponded purely linguistically, but rather that each item in each section tested the same construct in the original and the adaptation. It was also crucial that item equivalence was established. This does not necessarily mean that the item is translated as closely as possible in the target language, but rather that the item tests the same function in the target language as it does in the source language. For the NRDLs this implies that any test item in the Norwegian version should test a concept that is associated with a

certain developmental stage in the language development in children as it is in the original version. Furthermore, for the adaptation, the cultural aspect was not as crucial as the item and construct equivalence. Thus, as Hambleton (1994) presents, the most crucial guidelines for this adaptation would be the second, the eight and the fifteenth. Moreover, what Brislin (1986) says about being familiar with what the tool measures, what research the test is based on and what implications the results of the test were especially crucial for the adaptation of the NRDLs.

An example of item equivalence regarding the adaptation of the Norwegian NRDLs, is item 62 of section (G) in the comprehension part. Here, the English original reads 'The baby is pushed by the mother'. In the Norwegian version, this has been adapted to read 'Babyen blir trillet av moren', (The baby is wheeled by the mother). Here, the verb 'push' has not been translated by what is arguably its closest equivalent, 'dytte', but rather 'trille' (wheel). As the Norwegian form of the verb 'push' (dytte) has more negative connotations and is generally regarded as unacceptable to use in this context, a more acceptable verb has been used, 'wheel'. It also seems that although there are semantic differences between the verbs 'push' in English and 'trille' in Norwegian, for the context of this item together with the image that is presented to the child, both verbs seem to convey a relatively similar meaning (Vulchanova et al., 2019). Most importantly, the Norwegian item tests the same construct as the English version, namely the passive construction, and thus it appears that construct equivalence has been achieved.

Lastly, it should be noted that a back-translation of the Norwegian translation was conducted, and it was found to satisfy the requirements for sufficient adaptation, cultural appropriateness and construct and item equivalence.

5.4 The possible uses of the NRDLs

Generally, as discussed previously, the NRDLs is intended to be used with children between the ages of 2:00 and 7:05 to screen for possible speech or language deficits and can usually be used as a starting point for determining whether further screening or possible intervention is necessary in individuals. The NRDLs may also be used in scientific contexts, for instance in a comprehensive study of individuals in a certain age bracket or several age brackets and their language development, speech development, or the predicted developmental trajectory

for certain linguistic features, to mention some. Furthermore, there are certain other uses the NRDLs may have, such as studying SES and multilingual children, which will be discussed in more detail in the next sub-sections.

5.4.1 NRDLs and socio-economic factors

In a study published in 2013 as a result of the development of the English NRDLs, Letts, Edwards, Sinka, Schaefer and Gibbons examined the performance on the NRDLs with regards to the socio-economic status of the examined children. They found that the effect of socio-economic status was not as great as predicted, and that the effect found was more present in production than comprehension. Furthermore, one of the most central findings was that the relation between socio-economic status and vocabulary was more salient than that between socio-economic status and language more generally. The only significant difference was found between those whose parents had the education in the lower and upper quintiles. Group number 1, with individuals whose family had the lowest socio-economic status demonstrated a significant difference from the three other groups in the study in terms of production (Letts et al., 2013). The effect was most prominent in children under the age of 3:06 years who had mothers with the least education. Also, children in the age brackets between 4:06 and 5:06 years performed worse. Letts et al. (2013) conclude that this study is modest proof that lower socio-economic status affects children's result on the NRDLs, and that this effect is not found to the same degree in language production and language comprehension. The relationship between socio-economic status and performance on the test was not equal across the different age groups, and overall the size of the effects was modest (Letts et al., 2013). It seems that as long as children are in school or kindergarten full-time, the effects of the SES seem to decrease.

Letts et al. (2013) mention possible confounds that may affect the performance on the test, such as the fact that other skills than purely linguistic skills may be needed to perform well, such as executive functioning, reason, conceptual knowledge, attention, compliance, and working memory (Letts et al., 2013). Some evidence suggests that at least some of these can be related to SES (Ready & Reid, 2019; Brito, Piccolo & Noble, 2017). As for the effect of SES on performance on the NRDLs in Norwegian children, this is something that is yet to be studied, and SES was a data point that was not possible to measure in the present collection of the NRDLs data. However, this might be a point of future research on a Norwegian

population, as some research conducted with Norwegian children and teenagers suggest that there might be a link between SES and performance in school (Kjærnsli & Olsen, 2013).

5.4.2 Using the Multilingual Toolkit or the standard NRDLs?

The NRDLs comes with an instruction manual entitled Multilingual Toolkit, which details the process of administering the test on children that has a different L1. The idea is that the test administrator should speak the same language as the individual that is being tested, and then adapt the items in the NRDLs into the language of the children being screened (Letts & Sinka, 2011).

A question that arises is whether it is truly more beneficial to use the Multilingual Toolkit, or simply use the standard NRDLs for the language community in which the child resides, even though the child might not yet have achieved fluency in that language. As they necessarily attend kindergarten and school in the target language community it could be argued that it will be more beneficial to use that version of the NRDLs, rather than attempting to translate or use an adapted version for the individual's L1. Furthermore, some newer research by Vulchanova, Foyn, Nilsen and Sigmundsson (2014) has shown that children's competences in their L1 tend to correlate with their competences in other languages they know. In the research mentioned, the languages in question were Norwegian for the L1 and English for the L2, and it might be arguable if the effect was significant due to the fact that Norwegian and English are two relatively similar languages when it comes to language structure, argument patterns and vocabulary (Vulchanova et al., 2019). Nevertheless, further studies could determine if this effect is universal for acquisition of languages other than an individual's L1, or if there are transfer effects due to the similarity of the two languages in question (Carrasco-Ortíz et al., 2017).

Lastly, it should be noted that although a child might perform worse on the NRDLs than what the age of the individual would indicate according to the normed data, it does not necessarily indicate that the child's language development is delayed. There might simply be a need of more language exposure (Vulchanova et al., 2019) This especially holds for children who do not have the majority language of the community they are currently residing in as their L1.

5.4.3 Other possible uses and future research

Some other possible uses for the NRDLs outside children suspected to have certain linguistic developmental deficits, could be to use it with children who have cochlear implants and immigrant adults. At the time of writing, no extensive research on these topics have been carried out. Nevertheless, these are areas of interest that can be explored further in the future. For instance, it might be of interest to study the developmental trajectory of children that have cochlear implants, how their language develops after they have had it installed, and if the NRDLs can be used to determine if their language acquisition follows the projected trajectory.

As for the use of the NRDLs with adult immigrants, it might be of interest to study whether the NRDLs can be used as a tool for determining the language progress in immigrants who began acquiring Norwegian as adults. A possibility would be to use the NRDLs to test adults at two points in time and see if it is possible to use the NRDLs to determine their progress, and to determine which areas of language acquisition they are struggling with in order to implement the appropriate measures. The question remains whether certain parts or items of the test need to be changed or adjusted, as the NRDLs primarily is aimed at young children, and might thus appear too childish for i.e. adult males.

As mentioned in section 5.4.1, there is also the possibility to examine what effects SES might have on the performance on the NRDLs, as done with the population of English children during the standardisation of the NRDLs. Furthermore, as mentioned in section 5.2, the effect of gender on the performance need further research, and although it remains debatable whether separate norms for each gender are needed, it would at least be an aspect to consider for future work with the NRDLs.

Lastly, as Vulchanova et al. (2019) mention in the Norwegian manual, separate norms for children with other language backgrounds than Norwegian might also be of interest in the future. However, for the collection of the initial Norwegian norms, there simply were not enough children that had other language backgrounds than Norwegian that participated, and thus it was not possible to have separate norms for this group. Nevertheless, this might be something that could be explored in the future, and possibly by collecting data from a larger area in the rest of Norway in addition to Eastern Norway and Trøndelag.

6.0 Conclusion

The aim of this thesis was to be a descriptive account of the NRDLs and its possible uses, the theoretical background for its parts, as well as the process of adaptation and standardisation of the Norwegian NRDLs. It also aimed to account for some considerations regarding development and adaptation of standardised tests for primary use in clinical contexts.

The thesis has looked at different types of standardised tests, as well as different ways standardised tests are administered. Generally, standardised tests are useful tools for use in clinical contexts. Primarily, clinical tests such as NRDLs or TROG-2 can be used to identify possible developmental delays in children and to justify interventive measures to parents, teachers or guardians. Since most clinical tests tend to be well-tested and researched, it might be of interest to translate and adapt an existing test instrument into a different language instead of designing a new one. However, there are some considerations to pay attention to when adapting an existing instrument into a different language context. Brislin (1986) and Hambleton (1994) present a series of guidelines for development and adaptation of test instruments, and among these are concepts such as construct and item equivalence especially important. Moreover, familiarity with the structure, content and constructs of the test is also crucial for test adaptation.

The NRDLs is an example of a notable clinical test instrument that has gone through multiple revisions, as well as extensive evaluation and piloting. Moreover, it has been adapted for use in several other languages, such as Swedish and Norwegian. This thesis has looked at the process of standardising the Norwegian NRDLs, and some challenges that occurred both before and during data collection.

Currently, there are several possible uses for the NRDLs that remain unexplored, that might be of interest for research in the future. One of these issues are the relationship between SES and performance on the NRDLs in Norwegian-speaking children. Another possible area of further research is the subtle, but still persistent gender differences that seem to be present in Norwegian-speaking children, amongst others. Girls seem to have a subtle advantage over boys, especially in language production, and it might be of interest to study this difference further. Another possibility is to evaluate whether it would be useful to develop separate norms for each gender. Yet another area of interest would be to study multilingual children further using the NRDLs, both those that have Norwegian as an L1 in addition to other

languages, and those that do not have Norwegian as an L1 but are acquiring the language as an L2. Here, too, it could be useful to evaluate whether separate NRDLs norms should be developed for children with more than one L1, or even for children with other L1s than Norwegian.

At the time of writing, the Norwegian NRDLs has only been in print for a few months, and thus it has yet to be used widely in kindergartens, schools and other clinical contexts. Nevertheless, hopefully it can, like its predecessor, become a useful clinical tool for detecting language impairments and atypical language development in children, as well as possibly contributing to further research into language development and acquisition.

7.0 Literature

- Aguiar, N. & Taylor, M. (2015). Children's concepts of the social affordances of a virtual dog and a stuffed dog. *Cognitive Development*, 34, 16-27.
- Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26(2), 339-356.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Baillargeon, R. & DeVos, J. (1991). Object Permanence in Young Infants: Further Evidence. *Child Development*, 62(6), 1227-1246.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.
- Berk, L. (2012). *Infants and children: Prenatal through middle childhood* (7th ed.). Boston: Allyn & Bacon.
- Bishop, D. V. M. (2003). *Test for Reception of Grammar II. Manual*. Pearson Assessment: London.
- Bishop, D. V. M. (2009). *Test for Reception of Grammar II. Manual. Norsk versjon*. Pearson Assessment: London.
- Bricker, D. & Squires, J. (1999). *Ages and Stages Questionnaires: A parent-completed, child monitoring system*. Baltimore: Brooks Publishing Co.
- Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Newbury Park, CA: Sage.
- Brito, N., Piccolo, L. & Noble, K. (2017). Associations between cortical thickness and neurocognitive skills during childhood vary by family socioeconomic factors. *Brain and Cognition*, 116, 54.
- Bornstein, M., Haynes, O. M., Pascual, L., Painter, K. & Galperin, C. (1999). Play in two societies: Pervasiveness of process, specificity of structure. *Child Development*, 70, 317-331.

- Caldera, Y., Houston, A. & O'Brian, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine and neutral toys. *Child Development*, 60, 70–76.
- Caramazza, A. & Hillis, A. (1991). Lexical Organization of Nouns and Verbs in the Brain. *Nature*, 349(6312), 788.
- Carrasco-Ortíz, H., Herrera, A. V., Jackson-Maldonado, D., Ramírez, G. N. A., Pereyra, J. S. & Wicha, N. Y. (2017). The role of language similarity in processing second language morphosyntax: Evidence from ERPs. *International Journal of Psychophysiology*, 117, 91-110.
- Caselli, M. C., & Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L. & Weir, J. (1995). A Cross-Linguistic Study of Early Lexical Development. *Cognitive Development*, 10(2), 159-199.
- Chien, Y. & Wexler, K. (1990). Children's Knowledge of Locality Conditions in Binding as Evidence for the Modularity of Syntax and Pragmatics. *Language Acquisition*, 1(3), 225-295.
- Cronbach, L. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Crow, T. (1998). Sexual selection, timing and the descent of man: A theory of the genetic origins of language. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 17, 1079– 1114.
- Dabašinskienė, I. (2012). Gender Differences in Language Acquisition: A Case Study of Lithuanian Diminutives. *Journal of Baltic Studies*, 43(2), 177-196.
- Dale, P. S. & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*, 28(1), 125-127.
- Dick, F., Wulfeck, B., Krupa-Kwiatkowski, M. & Bates, E. (2004). The development of complex sentence interpretation in typically developing children compared with children with specific language impairments of early unilateral focal lesions. *Developmental Science*, 7(3), 360-377.
- Dunn, M. & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test - 4*. Circle Pines, MN: AGS.

- Edwards, S., Letts, C. & Sinka, I. (2011). *The New Reynell Developmental Language Scales. Manual*. London: GL Assessment.
- Edwards, S. & Varlokosta, S. (2007). Pronominal and anaphoric reference in agrammatism. *Journal of Neurolinguistics*, 20(6), 423-444.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M. & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30, 326-343.
- Espenakk, U. (Ed.). (2003). TRAS-håndbok. Stavanger: Høgskolen i Stavanger, Senter for leseforskning.
- Fava, L., Strauss, K., Valeri, G., D'Elia, L., Arima, S. & Vicari, S. (2011). The effectiveness of a cross-setting complementary staff-and parent-mediated early intensive behavioral intervention for young children with ASD. *Research in Autism Spectrum Disorders*, 5(4), 1479-1492.
- Fazio B.B., Naremore R.C. & Connell P.J (1996). Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech and Hearing Research*, 39, 611-624.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular.
- Fidler, A. & Rus, D. (2007). The role of linguistic complexity in early locative markings: a comparative study of Hungarian, Slovenian, and English. *Georgetown Papers in Theoretical Linguistics*, 7, 54-66.
- Gillam, R. B., Marquardt, T. P. & Martin, F. N. (Eds.) (2011). *Communication Sciences and Disorders* (2nd ed). Sudbury, MA: Jones and Bartlett.
- Gleitman, L. (1994). The structural sources of verb meanings. In *Language Acquisition*, P. Bloom (ed.), pp. 174-221. Cambridge, MA: MIT Press.
- Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A. & Trueswell, J. (2005). Hard Words. *Language Learning and Development*, 1(1), 23-64.
- Gleitman, H. & Gleitman L. (1997). What is a language made out of?. *Lingua*, 100, 29-55.

- Guasti, M. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.
- Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, *10*, 229–240.
- Hambleton, R. K., Merenda, P. F. & Spielberger, C. D. (2004). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In *Adapting educational and psychological tests for cross-cultural assessment* (pp. 15-50). Psychology Press.
- Hammarström, I. L., Kjellmer, L. & Hansson, K. (2017). *The New Reynell Developmental Language Scales. Manual. Svensk version*. Stockholm: Hogrefe.
- Hasselgård, H., Johansson, S. & Lysvåg, P. (2012). *English grammar: Theory and use* (2nd ed.). Oslo: Universitetsforlaget.
- Hide. 2019. In *Merriam-Webster.com*. Retrieved 12 May 2019, from <https://www.merriam-webster.com/dictionary/hide>.
- Jacobsen, M. & Bruun, O. (2000). *Human rights and Asian values: Contesting national identities and cultural representations in Asia* (Vol. No. 6, Democracy in Asia). Richmond: Curzon.
- Janson, H. & Squires, J. (2004). Parent completed developmental screening in a Norwegian population sample: A comparison with U.S. normative data. *Acta Paediatrica*, *93*, 1525–1529.
- Karmiloff, K. & Karmiloff-Smith, A. (2001). *Pathways to Language – From Fetus to Adolescent*. Cambridge, M/A London, England: Harvard University Press.
- Kaufman, A. S. & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test. Manual*. Pearson: Minneapolis.
- Kim, M., McGregor, K. & Thompson, C. (2000). Early Lexical Development in English- and Korean-Speaking Children: Language-General and Language-Specific Patterns. *Journal of Child Language*, *27*(2), 225-54.
- Kirk, C. & Vigeland, L. (2014). A Psychometric Review of Norm-Referenced Tests Used to Assess Phonological Error Patterns. *Language, Speech, and Hearing Services in Schools*, *45*(4), 365-377.

- Kjærnsli, M. & Olsen, R.V. (2013). PISA 2012 – sentrale funn. In *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012*. Oslo: Universitetsforlaget.
- Larsen, T. L. R. (2014). *Byggekløssar i barnespråk: Om tre norske born si tileigning av funksjonelle kategoriar* (Master's thesis, Norges teknisk-naturvitenskapelige universitet, Det humanistiske fakultet, Institutt for språk og litteratur).
- Letts, C., Edwards, S., Schaefer, B. & Sinka, I. (2014). The New Reynell Developmental Language Scales: Descriptive account and illustrative case study. *Child Language Teaching and Therapy*, 30(1), 103-116.
- Letts, C., Edwards, S., Sinka, I., Schaefer, B. & Gibbons, W. (2013). Socio-economic status and language acquisition: children's performance on the new Reynell Developmental Language Scales. *International journal of language & communication disorders*, 48(2), 131-143.
- Letts, C. & Sinka, I. (2011). *The New Reynell Developmental Language Scales. Multilingual Toolkit*. London: GL Assessment.
- Lyster, S. (2008). Norsk standardisering av TROG-2 (Test for reception of grammar-2); en presentasjon og en vurdering. *Norsk Tidsskrift for Logopedi*, 54(3), 25-26.
- Lyytinen, P., Poikkeus, A. M. & Laakso, M. L. (1997). *Language and symbolic play in toddlers. International Journal of Behavioral Development*, 21(2), 289–302.
- Marinis, T. & Chondrogianni, V. (2011). Comprehension of reflexives and pronouns in sequential bilingual children: Do they pattern similarly to L1 children, L2 adults, or children with specific language impairment? *Journal of Neurolinguistics*, 24(2), 202-212.
- Masterson, J., Druks, J. & Gallienne, D. (2008). Object and action picture naming in three- and five-year-old children. *Journal of Child Language*, 35(2), 373-402.
- McClure, K., Pine, J. M. & Lieven, E. V. M. (2006). Investigating the abstractness of children's early knowledge of argument structure. *Journal of Child Language*, 33, 693-720.
- McDonald, J. L. (2008). Grammaticality judgements in children: the role of age, working memory and phonological ability. *Journal of Child Language*, 35, 247-268.

- Mitrofanova, N. & Westergaard, M. (2018). Acquisition of locative utterances in Norwegian: Structure-building via lexical learning. *Journal of Child Language*, 45(4), 981-1005.
- Montgomery, J. & Evans, J. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 52(1), 269-288.
- Morrison, C. M., Hirsh, K. W. & Duggan, G. B. (2003). Age of acquisition, ageing and verb production: normative and experimental data. *Quarterly Journal of Experimental Psychology*, 56(4), 705-730.
- Phelps-Terasaki, D. & Phelps-Gunn, T. (2007). *Test of Pragmatic Language, Second Edition*. East Moline, IL: Linguisticsystems
- Piaget, J., Vonèche, J. & Gruber, H. (1995). *The essential Piaget* (2nd ed.). Northvale, N.J: J. Aronson.
- Protassova, E. & Voeikova, M. D. (2007). Diminutives in Russian at the Early Stages of Acquisition. In Dressler, W., & Savickienė, I. (Eds). *The Acquisition of diminutives: A cross-linguistic perspective* (Vol. V. 43, Language acquisition & language disorders). Amsterdam: John Benjamins Pub.
- Ready, D. & Reid, J. (2019). Children's executive function development and school socio-economic and racial/ethnic composition. *Early Childhood Research Quarterly*, 47, 457-471.
- Rice, M. & Wexler, K. (1996). Towards tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language and Hearing Research*, 39(6), 1239-1257.
- Rice, M. & Wexler, K. (2001). *Test of Early Grammatical Impairment: Examiner's manual*. San Antonio: Psychological Corporation.
- Røyneland, U. (2009). Dialects in Norway: Catching up with the rest of Europe?. *International Journal of the Sociology of Language*, (196-197), 7-30.
- Shawer, S. (2010). *Standardized assessment and test construction without anguish: The complete step-by-step guide to test design, administration, scoring, analysis, and interpretation (Education in a competitive and globalizing world)*. Hauppauge, NY: Nova Science.

- Sheridan, M., Frost, M. & Sharma, A. (1997). *From birth to five years: Children's developmental progress* (Rev. and updated [ed.] by Marion Frost and Ajay Sharma. ed.). London: Routledge.
- Siewierska, A. (2013). Passive Constructions. In Dryer, M. S. & Haspelmath, M. (Eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Sireci, S.G. & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing*, 20(2), 148–166.
- Squires, J., Bricker, D. & Potter, L. (1999). *The ASQ user's guide (2nd ed.)*. Baltimore: Paul H. Brookes.
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L. & Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC medical research methodology*, 10(1), 1.
- Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA : Harvard University Press.
- Tsai, H. A., McClelland, M. M., Pratt, C. & Squires, J. (2006). Adaptation of the 36-month Ages and Stages Questionnaire in Taiwan: Results from a preliminary study. *Journal of Early Intervention*, 28, 213–225.
- Valian, V. (2006). Young children's understanding of present and past tense. *Language, Learning and Development*, 2(4), 251-276.
- Van de Vijver, F. J. & Hambleton, R.K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89–99.
- Van de Vijver, F. J. & Poortinga, Y. H. (2004). Conceptual and methodological issues in adapting tests. In *Adapting educational and psychological tests for cross-cultural assessment* (pp. 51-76). Psychology Press.
- Van der Lely, H. & Stollwerck, L. (1997). Binding theory and specifically language impaired children. *Cognition*, 62, 245-290.
- Van Heerden, A., Hsiao, C., Matafwali, B., Louw, J. & Richter, L. (2017). Support for the feasibility of the ages and stages questionnaire as a developmental screening tool: A cross-sectional study of South African and Zambian children aged 2-60 months. *BMC Pediatrics*, 17(1), 55.

- Vulchanova, M., Foyn, C., Nilsen, R. & Sigmundsson, H. (2014). Links between phonological memory, first language competence and second language competence in 10-year-old children. *Learning and Individual Differences*, 35, 87-95.
- Vulchanova, M., Vulchanov, V. & Nilsen, R. (2019). *The New Reynell Developmental Language Scales. Manual. Norsk versjon*. Stockholm: Hogrefe.
- Warren, P. (2013). *Introducing Psycholinguistics*. Cambridge: Cambridge University Press.
- Yang, M. & Yu, R. (2008). The interviewer effect when there is an education gap with the respondent: Evidence from a survey on biotechnology in Taiwan. *Social Science Research*, 37(4), 1321-1331.
- Åfarli, T. A. (1989). Passive in Norwegian and in English. *Linguistic Inquiry*, 101-108.

Appendices

Appendix A

Consent form distributed to parents of children in kindergartens

Kjære foreldre/foresatte

Ditt barns barnehage hjelper oss i år med et prosjekt vi holder på med ved NTNU. Vår forskergruppe har som målsetning å kartlegge språkferdigheter hos norske barn i ulike aldersgrupper, og i den anledning trenger vi testpersoner med norsk som morsmål, alternativt to-/flerspråklige barn med norsk som ett av sine førstespråk. For å oppnå pålitelige og representative resultater, er vi avhengig av at et stort antall barn fra flere barnehager deltar i prosjektet, og i den forbindelse ber vi om tillatelse til å teste ditt barns morsmålsferdigheter. Vi anvender velutprøvde testmetoder som barn synes er artige og ikke oppfatter som skremmende. Testresultatene er det kun prosjektets medarbeidere som har tilgang til, og resultatene vil selvsagt være fullstendig anonymiserte før de brukes i offentligheten. Testingen skal foregå i barnehagen der barna testes individuelt. Prosjektet er meldt til NSD (Personvernombudet for forskning)..

Etter at prosjektet er gjennomført, vil alle navn, bakgrunnsopplysninger om barna og annet innsamlet materiale som kan bidra til å identifisere enkeltindivider, bli slettet.

Hvis du har spørsmål eller ønsker mer informasjon om prosjektet, må du svært gjerne kontakte oss.

Hvis du gir din tillatelse til at ditt barn kan delta i prosjektet, krysser du av i 'ja'-boksen nedenfor og innleverer skjemaet i utfylt stand med informasjon om barnets språkferdigheter til personalet i barnehagen så snart som mulig.

Samtykker du i at barnet ditt deltar, kan du selvsagt likevel ombestemme deg og reservere deg når som helst!

Med vennlig hilsen

Mila Vulchanova, Randi Alice Nilsen, Valentin Vulchanov

Kontakt: Mila Vulchanova, Professor
Språktilegnelses- og språkprosesseringslab, Institutt for språk og litteratur
NTNU
7491 TRONDHEIM
Tlf. 73596794, epost: mila.vulchanova@ntnu.no

SAMTYKKEERKLÆRING

Ja, jeg samtykker i at mitt barn deltar i forskningsprosjektet, og har krysset av for hans/hennes språkbakgrunn/-ferdigheter på det vedlagte skjemaet.

Barnets navn

Sted og dato

Foresattes underskrift

Opplysninger om barnets språkbakgrunn/-ferdighet

Hvis du godtar at barnet deltar i prosjektet, krysser du av for det alternativet nedenfor som passer best for ditt barn:

- Barnet har ingen spesielle kunnskaper i andre språk enn norsk.
- Barnet snakker ___ angi språk _____ bedre enn han/hun snakker norsk.
- Barnet snakker ___ angi språk _____ omtrent like godt som han/hun snakker norsk.
- Barnet snakker ___ angi språk _____, men han/hun snakker norsk bedre.
- Barnet har ingen hørsels-/språklige/motoriske problemer

Bruk gjerne kommentarfeltet under til flere opplysninger, f.eks. bakgrunnen for at barnet evt. er tospråklig og om nivået på språkene.

Si gjerne også noe om hvordan dere oppfatter barnets språklige og motoriske ferdigheter.

Barnets navn

Sted og dato

Foresattes underskrift

Appendix B

Consent form distributed to parents of children in schools

Kjære foreldre/foresatte

Ditt barns skole hjelper oss i år med et prosjekt vi holder på med ved NTNU. Vår forskergruppe har som målsetning å kartlegge språkferdigheter hos norske skoleelever i ulike aldersgrupper, og i den anledning trenger vi testpersoner med norsk som morsmål, alternativt to-/flerspråklige barn med norsk som ett av sine førstespråk. For å oppnå pålitelige og representative resultater, er vi avhengig av at store elevgrupper fra flere skoler deltar i prosjektet, og i den forbindelse ber vi om tillatelse til å teste ditt barns morsmålsferdigheter. Vi anvender velutprøvde testmetoder som barn syns er artige og ikke oppfatter som skremmende. Testresultatene er det kun prosjektets medarbeidere som har tilgang til, og resultatene vil selvsagt være fullstendig anonymiserte før de brukes i offentligheten. Testingen skal foregå i skoletiden og vil ikke overstige 2 timer totalt. Elevene testes individuelt. Resultatene kan senere brukes veiledende for lærere og læremateriellprodusenter om hvordan språkferdigheter utvikles, og de vil kunne bidra til å avdekke områder der enkeltelever måtte slite med (lære)vansker. Prosjektet er meldt til NSD (Personvernombudet for forskning).

Etter at prosjektet er gjennomført, vil alle navn, bakgrunnsopplysninger om barna og annet innsamlet materiale som kan bidra til å identifisere enkeltindivider, bli slettet.

Hvis du har spørsmål eller ønsker mer informasjon om prosjektet, må du svært gjerne kontakte oss.

Hvis du gir din tillatelse til at ditt barn kan delta i prosjektet, krysser du av i 'ja'-boksen på svarslippen nedenfor før den signeres og sendes til Mila Vulchanova på den oppgitte adressen, sammen med vedlagte skjema i utfylt stand med informasjon om barnets språkferdigheter. Svarkonvolutt m/porto følger vedlagt.

Samtykker du i at barnet ditt deltar, kan du selvsagt likevel ombestemme deg og reservere deg når som helst!

Med vennlig hilsen

Mila Vulchanova, Randi Alice Nilsen, Valentin Vulchanov

Kontakt: Mila Vulchanova, Professor
Språktilegnelses- og språkprosesseringslab, Institutt for språk og litteratur
NTNU
7491 TRONDHEIM
Tlf. 73596794, epost: mila.vulchanova@ntnu.no

SAMTYKKEERKLÆRING

Ja, jeg samtykker i at mitt barn deltar i forskningsprosjektet, og har krysset av for hans/hennes språkbakgrunn/-ferdigheter på det vedlagte skjemaet.

Barnets navn

Sted og dato

Foresattes underskrift

Opplysninger om barnets språkbakgrunn/-ferdighet

Hvis du godtar at barnet deltar i prosjektet, krysser du av for det alternativet nedenfor som passer best for ditt barn:

- Barnet har ingen spesielle kunnskaper i andre språk enn norsk.
- Barnet snakker ____ angi språk _____ bedre enn han/hun snakker norsk.
- Barnet snakker ____ angi språk _____ omtrent like godt som han/hun snakker norsk.
- Barnet snakker ____ angi språk _____, men han/hun snakker norsk bedre.
- Barnet har ingen hørsels-/språklige/motoriske problemer

Bruk gjerne kommentarfeltet under til flere opplysninger, f.eks. bakgrunnen for at barnet evt. er tospråklig og om nivået på språkene.

Si gjerne også noe om hvordan dere oppfatter barnets språklige og motoriske ferdigheter.

Barnets navn

Sted og dato

Foresattes underskrift

Appendix C

Images of the contents of the NRDLs



Image 1: Image taken from the NRDLs picture book, the production part, section (Cii), item 29.

COMPREHENSION SECTION Cii
(Verbs)



Image 2: Image taken from the NRDLs picture book, the comprehension part, section (Cii), item 26.

PRODUCTION SECTION Dii
(Sentence Building)



Copyright © 2011 GL Assessment

xvii

Image 3: Additional image cards, for use with production section (Dii).



Image 4: Objects used in the NRDLs. Pictured: Bed, table, chair, spoon, cup.

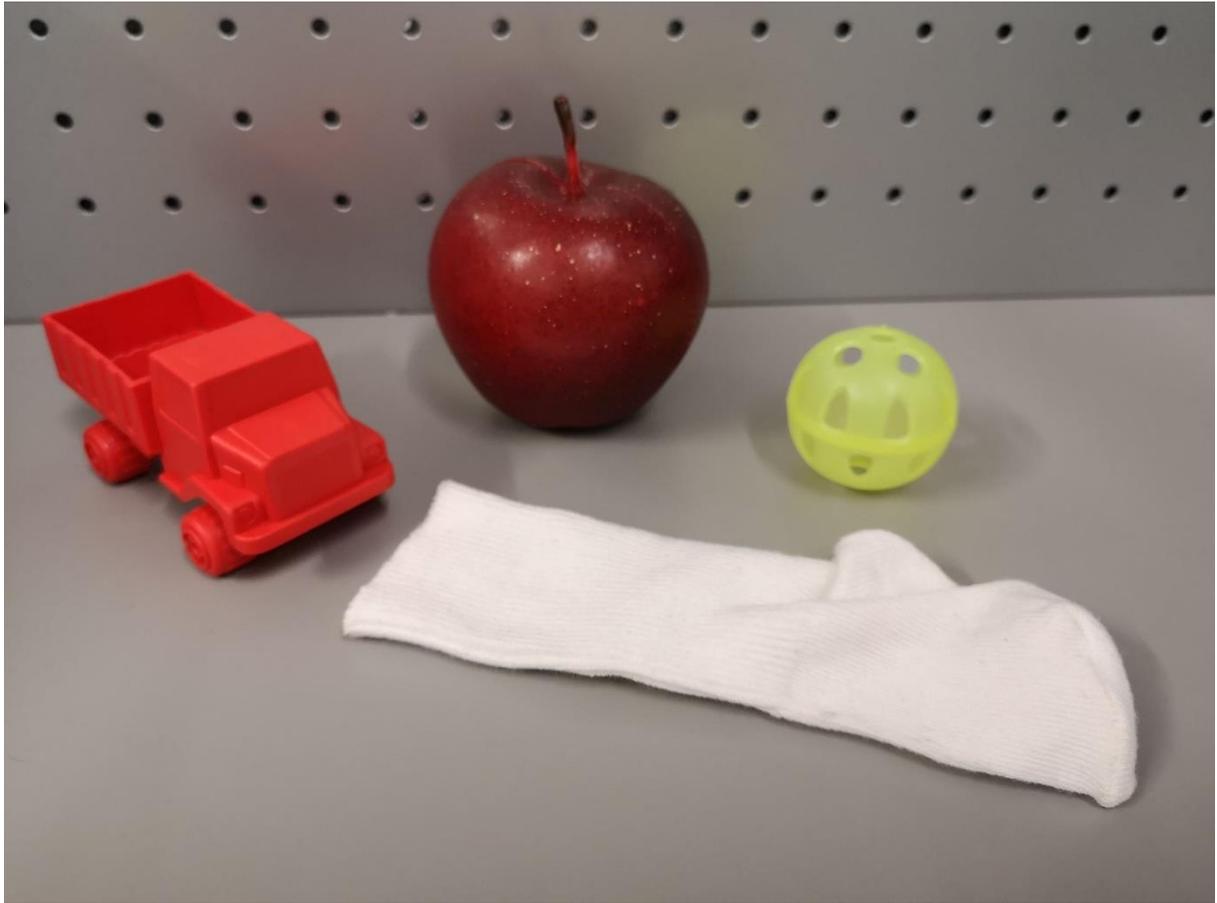


Image 5: Objects used in the NRDLs. Pictured: Truck, apple, ball, sock.



Image 6: Objects used with the NRDLs. Pictured: Duck, box, pencil, brush.



Image 7: Objects used with the NRDLS. Pictured: Rabbit, Monkey, Teddy.

Appendix D

Table 1.0

Number of children participating in the Norwegian standardisation of the NRDLs, distributed across age and gender

Age bracket	Boys	Girls	Children in total
3:00-3:05	7	9	16
3:06-3:11	14	28	42
4:00-4:05	18	18	36
4:06-4:11	26	20	46
5:00-5:05	14	11	25
5:06-5:11	11	14	25
6:00-6:05	10	8	18
6:06-6:11	7	4	11
7:00-7:10	13	10	23
Total	120	122	242

Appendix E

Relevance for the Teacher Education

As this master's thesis completes the five-year Teacher Education programme I am currently enrolled in, I will say a few words about how my work with this thesis is be relevant for the teaching profession.

Through my work with this master's thesis, I have primarily worked with data collection from children in the target age group for the NRDLs (2-7-year olds), which is much younger than the students I will primarily be teaching when I finish my education and start working. Nevertheless, the practice of working extensively with language testing and children has proved to be an invaluable experience, which surely will be relevant in my future career as a teacher. During the course of the work on this thesis I have mostly studied how Norwegian children acquire their L1, however, this might also be transferable knowledge to how older teenagers acquire one or several L2s. As my subjects of specialisation are English and French, language acquisition is undoubtedly one of the areas of great interest and usefulness for my future teaching profession. Understanding how the process of first language acquisition in children functions will necessarily be transferable to L2 acquisition in older children and is therefore particularly relevant.

Through my work with this thesis I have also gained valuable insight into clinical testing and tools for detecting possible atypical language development. Both by administering the NRDLs, but also TROG-2, which is a tool intended both for young children, but also teenagers and adults. The experience with administering these types of language tests is something that have given me a lot of confidence in test administration and clinical testing, which might be something I would like to pursue further in the future.

Lastly, working with children in this type of setting has not only been both challenging and informative, but it has also been a great joy. Although the age group is quite different, it has honestly renewed my confidence in the fact that working with young people is something I truly want to do in the future.

