Tobias Haug

# Adaptation and Evaluation of a German Sign Language Test

Hamburg University Press

*Video embedded*

Embedded in this PDF file you will find a video (MP4-formatted) concerning “Examples of the Revised Version of the DGS Receptive Skills Test for Main Study” (see pp. 133–139).

In order to start the video, click on the „Play“-icon below.

If there are any problems in connection with playing the video, it is recommended to open the PDF file with Acrobat Reader X (http://get.adobe.com/de/reader/, free download).

Tobias Haug

Adaptation and Evaluation of a German Sign Language Test

Tobias Haug

# Adaptation and Evaluation of a German Sign Language Test

## A Computer-Based Receptive Skills Test for Deaf Children Ages 4–8 Years Old

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg
Carl von Ossietzky

Bibliographic information published by the *Deutsche Nationalbibliothek* (German National Library).
The *Deutsche Nationalbibliothek* lists this publication in the *Deutsche Nationalbibliografie*; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.
The online version is available online for free on the website of Hamburg University Press (open access). The *Deutsche Nationalbibliothek* stores this online publication on its Archive Server. The Archive Server is part of the deposit system for long-term availability of digital publications.
Available open access in the Internet at:
Hamburg University Press – http://hup.sub.uni-hamburg.de
PURL: http://hup.sub.uni-hamburg.de/purl/HamburgUP_Haug_Adaption
Archive Server of the *Deutsche Nationalbibliothek* – http://deposit.d-nb.de
ISBN 978-3-937816-79-1 (printed version)

Printing house: Elbe-Werkstätten GmbH, Hamburg, Germany
http://www.ew-gmbh.de
Cover design: Benjamin Guzinski, Hamburg

# Acknowledgment

First of all, I would like to thank my two academic advisors, Prof. Dr. Rolf Schulmeister from Hamburg University and Prof. Dr. Bencie Woll from University College London, for their guidance, discussions, feedback and continued support via phone, mail, and personal meetings during the different stages of my dissertation over the last seven years.

Additionally, I would like to thank those people without whose continued support I could not have made it: Rosalind Herman at City University London, for her input over the last couple of years and for spending a lot of time discussing the adaptation of the BSL Receptive Skills Test to DGS with me; most importantly, Rosalind Herman, Sallie Holmes, and Bencie Woll, who gave me permission to use the BSL Receptive Skills Test as a template for test adaptation; Dr. Penny Boyes Braem, who provided useful input throughout the process of writing and re-writing my dissertation before finally submitting it; and Prof. Dr. Manfred Hintermair from the University of Education in Heidelberg, who gave me encouragement and support via mail, phone, and personal meetings.

Furthermore, I want to express my gratitude to the people who helped me adapt the test: my programmer Michael Metzger for his very creative solutions while programming the testing interface for the adapted DGS Receptive Skills Test, as well as Knut Weinmeister from *Gebärdenwerk* in Hamburg, who was kind enough to function as a DGS model in the test. He did an outstanding job and the children who participated in this study loved him. Georg Eberhard and Jana Schwager from the sign language school *Sehen-Verstehen* in Heidelberg were very helpful in collecting data and advising me regarding DGS issues in the process of adapting the test to DGS. Margit Hillenmeyer from the GIB (Deaf Institute of Bavaria), Nürnberg, was very supportive while collecting data in the Pilot 1 stage and provided very useful input and offered suggestions regarding linguistic issues. A big "thank you" to Jovita Lengen (formerly GS-Media, Zurich), who was very skilled in changing and revising the original foils from the BSL test to cultural-appropriate foils for the German test version. I also thank Sabine Fries for collecting data, and Beate Krausmann and Sylvia Wolff, Humboldt University of Berlin, for their support and input, as well as coworkers from Hamburg University, Institute for German Sign Language: Stefan Goldschmidt, Thomas Hanke, Torsten Herwig, Thimo Kleyboldt, Gabriele Langer, Christiane Metzger, Tobias Meyer-Jansson, and Siegmund Prillwitz. Most of all, thanks goes to all the Deaf informants, both children and adults, who participated in the pilot or main studies.

Thanks goes to all the teachers and children at the participating schools; without them it would never have been possible to conduct the test.

In the difficult stage of statistical analysis, Jürg Schwarz and Esther Meier, University of Zurich, provided guidance and were very patient discussing statistical issues with me over the past year. Thanks also goes to Wolfgang Mann from City University London, who was very helpful in discussing parts of my dissertation. Chloe Marshall from City University London did a wonderful job of proofreading my dissertation; not only did she take a look at my L2-English, she also provided very useful comments and suggestions in the final stage.

I also would like to thank my employer, the University of Applied Sciences for Special Needs Education (HfH) in Zurich, for providing financial support to publish my dissertation at Hamburg University Press.

Last but not least, I can never express enough thanks to all those who have helped me with this work over the last few years; it is unfortunately not possible to mention them all here.

# Table of Contents

## List of Tables

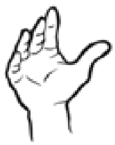
## List of Figures

## Abbreviations of Sign Languages

| | |
|---|---|
| ASL | American Sign Language |
| Auslan | Australian Sign Language |
| BSL | British Sign Language |
| DGS | Deutsche Gebärdensprache (German Sign Language) |
| LIS | Lingua Italiane dei Segni (Italian Sign Language) |
| LSB | Lingua de Sinais Brasileiros (Brazilian Sign Language) |
| LSF | Langue des Signes Française (French Sign Language) |
| NGT | Nederlandse Gebarentaal (Sign Language of the Netherlands) |

# Conventions for Glosses

| *Gloss* | *Meaning* |
|---|---|
| HOUSE | Sign (BSL, or any other sign language) with an equivalent meaning in English is capitalized. |
| HAUS (house) | Glosses in German always stand for sign of German Sign Language (DGS), and the English meaning is always added in parentheses. This is in order to differentiate between DGS signs and signs used in any other sign language. |
| HOUSE++ | The symbol "++" is added to a sign to indicate plurality (e.g., *two houses*). |
| CL | Indicates the use of a whole entity classifier. |
| BOY-left | The lower case "left" indicates additional information about the referent's location in sign space. |
| (BOY) | The referent is not mentioned explicitly. |
| CAR-ROW-ROW-ROW | Describing the meaning of a sign often requires the use of more than one word. Therefore, the hyphen between two or more glosses indicates that only one sign is being referred to (e.g., *rows of cars*). |
| Flat-B handshape |  B-handshape from the manual alphabet; it is the flat handshape, palm pointed downward, and, for example, indicates a classifier for *car* in BSL and DGS. |
| Upright-1 handshape |  Index finger extended from fist, for example, whole entity classifier representing a *person* in DGS. |
| Upright-3 handshape |  Thumb, index, and middle finger extended from fist, palm pointing sideward, for example, a classifier in ASL representing a *car*. |

| *Gloss* | *Meaning* | |
|---|---|---|
| Upright-4 handshape |  | All fingers except thumb extended and spread, palm pointing sideward, for example, classifier in BSL or DGS representing a *queue*. |
| 5-handshape |  | All fingers extended, palm pointing sideward, for example, handshape for the DGS sign HASSEN (to hate). |
| 5-Clawed handshape |  | Basis is the 5-handshape (see above), all fingers extended, spread and loosely curved, for example, in DGS following a nominal sign HAUS (house) and then HAUS-located. |
| V-Bent handshape |  | Index and middle fingers bent, palm pointing downward for example, a classifier in DGS for a four-legged animal sitting. |

(Beecken, Keller, Prillwitz, & Zienert, 1999; Sutton-Spence & Woll, 1999)

# 1 Introduction

## 1.1 Background

### 1.1.1 Deaf People as Bilinguals

Many Deaf[1] people in developed countries can be defined as bilinguals, using both a sign language and the majority language in written and/or spoken form in their everyday lives (Grosjean, 2008). Competency in the two languages can vary widely, depending, among other factors, on the age at which Deaf individuals are first exposed to a(n) (accessible) first language (L1) and second language (L2) (Mayberry & Lock, 2003). Abilities in a sign language can range on a continuum from Deaf children acquiring a sign language as a first language from their native signing Deaf parents, to Deaf children of hearing parents acquiring a sign language only when they enter school. In particular, this latter group of Deaf children who are born Deaf and have hearing parents might have delayed first language acquisition, and they comprise a special population where language is a crucial variable.

Deaf children who acquire a sign language as their first language from their Deaf parents constitute only about 5% of the population of Deaf children (Mitchell & Karchmer, 2004). For the remaining 95% who come from hearing families, acquiring a language is often a great challenge (Marschark, 2002). The majority of Deaf children of non-signing hearing parents do not have full access to a sign language until they have passed the most critical early ages of language acquisition. Language development can differ between the groups of, on the one hand, Deaf children with hearing parents

[1] It is a widely recognized convention to use the upper case *Deaf* for describing members of the linguistic community of sign language users, and in contrast, to use the lower case *deaf* when describing the audiological state of a hearing impairment (e.g., Morgan & Woll, 2002a). However, when referring to children it does not make sense to draw this distinction since it is not clear whether, for example, a 4-year-old deaf child of hearing parents is a member of this community of sign language users or not. But within the scope of this study, out of respect to the members of the community of sign language users, the convention of the upper case *Deaf* will be used.

and, on the other hand, children who receive native language input (Deaf children of Deaf parents, and hearing children of hearing parents).

With the technological advances of multichannel cochlear implants, the spoken language development of implanted children has improved (Blamey, 2003). Implanted children generally acquire speech in the same order as their hearing peers, and acquire it faster than children with a hearing-aid. However, the process is still slower than for hearing children acquiring speech, and "unintelligible speech remains the norm" (Marschark, 2002, p. 3; Marschark, Lang, & Albertini, 2002). Thus, implanted children generally lag behind in spoken language development relative to their hearing peers (Blamey, 2003; Marschark & Spencer, 2006). Early and meaningful access to a language provides Deaf children the opportunity for early language acquisition, which further results in long-lasting advantages in other domains of their development (Woll, 1998). It is also important to evaluate and monitor the sign language development of Deaf children from hearing families, that is, those children who have access to (sign) language models in early or later intervention programs.

Internationally, the education of Deaf children has changed over the past decades with the emergence of bilingual and bicultural programs in the US (e.g., Mahshie, 1995; Nover, 2005) and in several European countries (e.g., Germany: Günther, 1999; Günther & Schäfke, 2004; Austria: Krausneker, 2004; Denmark: Lewis, 1995). These programs use a sign language as the language of instruction for Deaf children, and in most cases, as a means upon which to build the knowledge of the written (and spoken) forms of the majority language as a second language (L2). In some countries, research has been conducted to evaluate the effectiveness of such programs and/or to investigate the relationship between a sign language as the L1 and the literacy skills of the majority language as the L2 (e.g., US: Hoffmeister, 2000; Prinz, 2002; Strong & Prinz, 1997, 2000; Germany: Mann, 2008; Switzerland: Niederberger, 2004, 2008). The results suggest a positive correlation between sign language skills and written skills in the majority language.

Plaza-Pust and Morales-López (2008) list some of the shortcomings of existing bilingual programs: "[T]he status assigned to the different languages and communication systems, teacher training, the materials used and assessment methods available strike us in their potential negative effects concerning the eventual outcomes" (p. 350). Relevant for the present study is the fact that a range of *evaluation procedures*, including sign language tests, are needed for bilingual programs to evaluate language development. The need for sign language tests in schools for the Deaf has been surveyed and confirmed in different countries (Switzerland: Audeoud & Haug, 2008; Germany: Haug & Hintermair, 2003; UK: Herman, 1998; US: Mann & Prinz, 2006).

### 1.1.2 Current Bilingual Deaf Education in Germany

The present study is an adaptation of an existing sign language test to be used for German Sign Language (DGS; *Deutsche Gebärdensprache*), and as such, an overview of current German educational practices for Deaf children would be pertinent. In 1992, the first pilot bilingual class was introduced at the school for the Deaf in Hamburg. This first trial class was scientifically evaluated (Günther, 1999; Günther & Schäfke, 2004) and was followed by a second bilingual trial in 2001 at the school for the Deaf in Berlin (Günther & Hennies, in press). At the present time, bilingual methods have become more accepted in Germany but they still constitute a minority of the educational approaches for Deaf children actually being employed in this country (Günther, Hennies, & Hintermair, 2009). Deaf children are fewer in number compared to hard-of-hearing children, and for their primary and secondary school education, the vast majority are educated in special schools and sometimes together with hard-of-hearing students and/or children with central auditory processing disorder (Günther et al., 2009). The official proportion of children with hearing loss that are mainstreamed is about 20%, and 90% of those children are hard-of-hearing (Günther et al., 2009). Concerning the modes of communication used in German schools for the Deaf (Große, 2003), the majority of Deaf institutions (90%) have the mastery of spoken language as their primary goal although manual means of communication are included to some extent in about 60% of the classes. These manual means range from "the use of the Phoneme Transmitting Manual System, to use of the manual alphabet, occasional use of signed German or German Sign Language through to a full, bilingual approach" (Günther et al., 2009, p. 183). This picture has to a lesser extent been confirmed by a survey of the need for sign language tests, in which respondents ($N = 203$) from 33 institutions (42% of 78 contacted) replied that some form of signing, ranging from LBG[2] (Signed German) to DGS, is used in their institution (Haug & Hintermair, 2003).

In many countries, the sign language evaluation carried out in preschools and primary schools is far from satisfactory. Singleton and Supalla (2003) point out that in practice, many schools in the US use informal descriptive evaluations of the Deaf children's signing skills, but these "assessment approaches are inadequate because they introduce multiple threats to the reliability and validity of the assessment results" (p. 289). The situation in Germany is no different. Of 203 returned questionnaires surveying the need for a DGS test, only 23 persons (from 9 institutions) replied that sign language skills are evaluated in their institution on a regular basis. As in the US, evaluation procedures are mostly informal, such as observations in class or video analysis. This is due to the absence of any standardized DGS measure (Haug & Hintermair, 2003). Because testing and monitoring the DGS development of Deaf children, particularly at an early age, is of great importance, there is a clear need for a sign language test that measures a range of linguistic devices which are important for language acquisition from the age of 3 onward.

### 1.1.3 Two General Approaches for Constructing Sign Language Tests

A reliable and valid test of DGS should be able to test specific aspects of language development (DGS) of Deaf children. A possible approach for researchers or test developers who want to design a sign language test that compares the individual performance of a child to his/her peers is to identify target structures that are acquired within a certain age span as reported in studies on sign language acquisition. However, the absence of almost any DGS acquisition studies (with the exception of Hänel, 2003), makes the development or adaptation based on existing tests difficult, if not impossible.

---

[2] *Lautsprachbegleitende Gebärden* (LBG) is not a language but a communication system primarily used in school settings to teach the structure of the German language in the visual modality (Wisch, 1990). In its pure form, LBG has a one-to-one representation of German lexical units in signs, and as such, resembles signing systems used in American schools such as "Signing Exact English". Sign language-specific structures, such as the use of space to encode verb agreement, are not utilized in LBG.

Sign language tests have been developed for other sign languages, specifically for American Sign Language (ASL) and British Sign Language (BSL), both of which, compared to DGS and many other sign languages, are better documented with regards to linguistic structure and acquisition (Haug, 2008a). These tests have been developed for different purposes; for example, for use in larger research projects such as evaluating bilingual programs (e.g., Test of ASL by Strong & Prinz, 1997, 2000)[3]. As most of these tests are still under development, one of their common weaknesses is their reported psychometric properties. Except for the BSL Receptive Skills Test (Herman, Holmes, & Woll, 1999) and the BSL Narrative Skills Test (Herman et al., 2004), there are few commercially available tests of sign languages.

In the absence of available research on their own sign languages, test developers in other countries, to a certain degree, have had to rely on (1) available research on the structure and acquisition of ASL or other better-researched sign languages (Fehrmann, Huber, Jäger, Sieprath, & Werth, 1995a, 1995b), or (2) tests developed for ASL or BSL, which they have used as a basis for the development of their own tests (e.g., Dubuisson, Parisot, & Vercaingne-Ménard, 2008).

Studies on test development for BSL (Herman et al., 1999) and ASL (Strong & Prinz, 1997, 2000), as well as studies on the acquisition of DGS and other sign languages, together with the available linguistic research of DGS structures, are the basis for the conceptual framework of the present study. In addition, this study will give a detailed description of the methodologies used to develop or to adapt a test; information which has been lacking in existing research, but which raises important issues for both research and practice, and not only for DGS but for the adaptation of tests between other sign languages as well.

## 1.2 Statement of the Problem

### 1.2.1 Need for a Sign Language Test of DGS

As discussed in the previous section, despite the current need for reliable and valid test instruments in different countries in order to monitor the

[3] Different sign language tests will be reviewed in Chapter 2, "Literature Review."

sign language acquisition of Deaf children (Maller, Singleton, S. Supalla, & Wix, 1999), very few tests that offer strong evidence of their psychometric properties are commercially available. A DGS test focusing on the linguistic structures acquired in preschool- and school-aged children (4–8 years old) is thus urgently needed.

### 1.2.2 Focus on Issues in Test Development

While issues involving test development have been addressed within the framework of a study as an instrument to investigate another research question, some studies address test development as their sole issue and focus (Herman, 2002; Hermans, Knoors, & Verhoeven, 2010). The issue of test adaptation from one sign language to another has been addressed to a certain degree in some studies (Johnston, 2004; Schembri et al., 2004). Only one review paper focuses on the sole issues in test adaptation (Haug & Mann, 2008). However, no previously conducted empirical study has focused explicitly on the linguistic, cultural, and methodological issues involved in the process of adapting a source language test to a target language test.

### 1.2.3 Test Adaptation

Considering the state of research in this field, test adaptation is a practical approach which offers the possibility of using an available template of appropriate test stimuli materials, together with the methodological and theoretical advantages of producing a test based on a reliable and valid test instrument. Using a sign language test which has been standardized and which has sound psychometric properties as a template for adaptation thus provides a starting point for tests of sign languages that have been less documented, such as DGS. However, it must be remembered that validity and reliability cannot be transferred to the adapted test; these need to be established anew for the adapted test. In terms of theory, adaptation offers both fertile ground for the discussion of cross-linguistic differences between sign languages and the opportunity to build up an adaptation model to be used for future test adaptations. Specifically, a thorough review and an analysis of cross-sign language studies can contribute to a hypothesis-building approach in test adaptation. In the long run, test adaptation offers the possibility for comparative studies in sign language acquisition based on the data obtained during testing in the source and target languages.

In sum, test adaptation is not only a practical approach, but it can also contribute to methodological and theoretical issues in the field of sign language testing.

## 1.3 Test Adaptation Approach Used in this Study

For the purpose of this study, the standardized BSL Receptive Skills Test (Herman et al., 1999) was used as a template for adaptation to DGS. Several different issues will be addressed: the need for a valid and reliable DGS test to be used in schools; the methodological reasons for choosing to begin by adapting an existing test; the state of research in DGS; and theoretical considerations relevant to the research questions.

### 1.3.1 Justification

The adaptation of the BSL Receptive Skills Test to DGS is important for a number of practical, methodological, and theoretical reasons. First and foremost, this study contributes to the adaptation and development of a test of DGS, which the survey (Haug & Hintermair, 2003) indicated to be of great importance to the schools. The findings will also contribute to a better understanding of the acquisition of DGS.

Methodologically, the present study focuses on language comprehension using a computer-based test, which not only applies new technology to sign language testing but also allows the specific needs of the target group to be met. Benefits to using a computer-based instrument for testing DGS development in Deaf children aged 4–8 include its standardized format and the fact that test results are automatically saved. These two features increase test reliability.

From a theoretical perspective, the present study makes a novel contribution to the field by examining linguistic, cultural, and methodological issues in the process of the adaptation from the source language test to the target language tests, and by providing explanations/models for future test adaptation. Important to the theory is the building up of hypotheses that are based on acquisition studies of other sign languages, studies on DGS, and cross-linguistic studies; factors that can be reformulated as a model to enhance the validity of the adapted test.

### 1.3.2 Research Questions Concerning the DGS Test

The present study is an adaptation of the BSL Receptive Skills Test to DGS, along with theoretical considerations of the linguistic and cultural aspects, and the psychometric properties of such an adaptation. The study addresses the following specific research questions:

1. Does the adapted DGS test provide evidence of having sound psychometric properties?
   1.1 Item analysis: Does the adapted DGS test show evidence of item facility and discrimination index?
   1.2 Fit of newly developed items: How do the newly developed items fit into the adapted test?
   1.3 Distractor analysis: Does the distractor analysis show evidence of the effectiveness of the distractors?
   1.4 Does the test show evidence of homogeneity?
   1.5 Does the test show evidence of reliability (e.g., Cronbach's alpha)?
   1.6 Does the test offer evidence of relations to an external variable (e.g., teachers' ratings of the children's sign language skills)?
   1.7 Does the test show evidence of content validity?[4]
2. What are the relationships between the Deaf children's raw scores and other variables (gender, age of sign language exposure, parental hearing status, chronological age)?
   2.1 Does the gender of the children have an impact on their test performance?
   2.2 Does the age of sign language exposure have an impact on children's test performance?
   2.3 Does parental hearing status have an impact on children's test performance?
   2.4 Does chronological age (in the subgroups of Deaf children of Deaf parents, and Deaf children of hearing parents) have an impact on children's test performance?

[4] This research question is more a theoretical/review-based question than an empirically-based question in this study and will be investigated in Chapter 5, "Discussion".

### 1.3.3 Methodological and Theoretical Questions and Issues

This research study investigates linguistic, cultural, and methodological issues in the adaptation of the BSL Receptive Skills Test to DGS. Recognizing the current state of research on DGS, the present study differentiates itself from other studies of test adaptation (Johnston, 2004; Schembri et al., 2002) by focusing on methodological and theoretical issues in test adaptation.

Based on these considerations, and drawing on evidence from studies of sign language acquisition, DGS, and cross-linguistic structures to adapt the BSL test to DGS in a first step, the argument will be made that test adaptation is not a straightforward procedure.

For practical reasons that are discussed in Chapter 3 ("Methodology"), the adapted DGS test will not be standardized within the frame of this study.

## 1.4 Scope of the Dissertation

In the next chapter, literature relevant to the present study will be reviewed and analyzed. Chapter 3 then describes the research designs and methods employed to collect, score, and analyze the data, as well as the test situation and criteria for selecting participants. The results of the study will be presented in Chapter 4, followed by a discussion in Chapter 5, where the findings will be summarized and discussed in relation to methodological, theoretical, and practical considerations.

# 2 Literature Review

The following chapter consists of seven main sections: (1) an overview of issues and terms in language testing followed by (2) different models for transferring tests across cultures and languages, and (3) an overview of test adaptation for spoken and sign languages. The fourth section (4) deals with a review of existing sign languages tests, followed by (5) a review of studies of sign language acquisition and cross-linguistic differences. The sixth section (6) links the previous issues of sign language acquisition and cross-linguistic differences to the adaptation of the DGS test, and the chapter concludes with (7) a summary and the implications of the state of knowledge of this field for the present study.

## 2.1 Issues in Language Testing

In our daily lives we are constantly confronted with testing. This starts in early childhood and continues throughout adulthood, with, for example, developmental screening or diagnostic tests in infancy, and tests to measure academic progress in primary school, school placement exams, final exams at university, driving exams, and testing procedures in connection with job applications. Evaluation and testing procedures are part of our social life and the culture we live in (e.g., Bartram, 1990; Fulcher & Davidson, 2007; McNamara, 2000).

In the following section, an overview of different issues in the development of (language) tests will be provided, starting with basic concepts that are relevant for testing and evaluation, and followed by important issues in language test development, including tests purposes and methods.

### 2.1.1 Basic Concepts in Language Testing

Two important concepts in language testing are *criterion-referenced* and *norm-referenced* tests. These are types of tests used in the evaluation process. In criterion-referenced tests, the candidate's score is not compared to the performance of a normative group, but rather, to a predefined criterion that needs to be achieved. This could be, for example, the final exam of a course, where the criteria of the course (knowledge, set of skills) are a set of clearly defined objectives which should be achieved by the end of the course and are therefore independent of the performance of others (Brown, 2004; Brown & Hudson, 2002; Davies et al., 1999). Norm-referenced tests are tests where the candidates' scores are interpreted with reference to the performance of other candidates who formed the normative group (Brown, 2004; Brown & Hudson, 2002; Davies et al., 1999). For example, the scores of a child on a norm-referenced and standardized test for language development can be interpreted to compare that child's performance to his/her peers who belong in the same age group and share the same characteristics, in order to make inferences about the child's language development.

Another important concept is the term *construct*:

> [A construct is a trait or behavior] that a test is intended to measure. A construct can be defined as an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores. A construct is generally defined in terms of a theory; in case of language, a theory of language. A test, then, represents an operationalisation of the theory. Construct validation involves an investigation of what a test actually measures and attempts to explain the construct. (Davies et al., 1999, p. 31)

The definition of a construct is based on a theoretical and abstract level (and also in broader terms) within the frame of a specific theory. For example, a construct relating to language could be defined as "fluency in a language" or "vocabulary knowledge". A construct definition is always subject to construct validation, that is, it requires "a construct theory upon which hypotheses can be developed and against which evidence can be evaluated" (Chapelle, 1999, p. 263). These basic definitions and terms are important for understanding the adaptation of the DGS Receptive Skills Test.

The term *(linguistic) competence* will be used in the present study in the Chomskian sense as the knowledge of a formal linguistic system at the level of grammar, as opposed to the application of this knowledge in language performance or actual language use (e.g., Brown & Hudson, 2002;

Davies et al., 1999). Within the context of language testing, an increased concern with actual language use rather than only its mental representation has led to a broadening of the definition of linguistic competence to also include communicative competence. In recent decades, for example, the definition has expanded to include discourse and pragmatic knowledge (e.g., Bachman, 1990; Canale & Swain, 1980; Hymes, 1972).

The following sections address issues such as test purposes and methods, which are important aspects in test development and adaptation.

### 2.1.2 The Goals of Language Testing in Children

A common goal of language testing in children is to see if a child's language development is following the expected course. To reach this goal, "a child's language skills are compared to the skills of same-age peers" (Johnston, 2007, p. 1). Another goal of language testing might be to describe the child's current language abilities so that "language therapy and school programming can be individualized" (Johnston, 2007, pp. 1–2). A third reason for language testing would be the "measurement of progress, either for an individual child, or an educational or therapeutic program" (Johnston, 2007, p. 2).

Wiig and Secord (forthcoming) explain different steps in the process, from the initial detection of children "at risk", to the design and delivery of intervention. In the first step, a language screening (or a screening for other developmental skills) is administered. A screening is defined "as a systematic procedure to select individuals from a given population at risk or an impairment" (Law, Boyle, Harris, Harkness, & Nye, 1998, p. iii). Screenings are usually administered to a large number of children, and are normally easy to administer and score. Criteria "for success or failure to pass age- or grade-level expectations may be designed for group or individual administration" (Wiig & Secord, forthcoming, Chapter 2, p. 8). Children who perform below a given criterion may be referred for in-depth diagnostic testing or classroom intervention (Wiig & Secord, forthcoming). Another path for identifying children "at risk" in order to refer them for diagnostic testing is the use of behavioral observations and rating scales. Rating scales "provide a standardized method for collecting information about a child or adolescent's strengths and weaknesses in broad areas such as listening, speaking, reading, and writing" (Wiig & Secord, forthcoming, Chapter 2, p. 9).

When children have been identified as having a potential language disorder, they may be referred for diagnostic testing. Diagnostic testing "focuses on critical dimensions or skills that define a disorder or syndrome" (Wiig & Secord, forthcoming, Chapter 2, p. 10). The objective of diagnostic testing is to clarify whether a child has a language disorder in a certain domain (McCartney, 1993) and whether she/he is therefore eligible for special provisions in the form of services or intervention. Once the presence of a disorder has been identified, its degree and possible associated deficits will be determined (Wiig & Secord, forthcoming). Once the disorder has been identified, intervention for the child will be planned. Clinical testing of language should provide accurate information, "upon which intervention can be based as quickly as possible in order to fulfill a useful function" (McCartney, 1993, p. 35).

The chosen purpose of a test consequently also determines the content of the test items.

### 2.1.3 Language Testing Methods for Children

Johnston (2007) categorizes language testing methods for children along two dimensions: (1) the nature of the sample, that is, elicited (paper-and-pencil) vs. spontaneous language sample (performance-based); and (2) the nature of reference, that is, norm- or criterion-referenced tests. Paper-and-pencil tests (also including computer-based tests) can have a fixed response format, for example, a multiple-choice format; whereas performance-based tests take place within a communicative context and tend to focus on speaking and writing (Davies et al., 1999; Fulcher & Davidson, 2007; McNamara, 2000).

Johnston (2007) also discusses in this context the advantages and disadvantages of specific language testing methods to be used with children. A major problem with norm-referenced tests that use elicited language behavior is, (1) that these tests are not sensitive enough to measure an individual child's language progress since they are constructed to yield stable scores over time, and (2) that the language "elicited by the items is non-communicative and decontextualized" (Johnston, 2007, p. 2), and differs from children's everyday language use. Johnston (2007) observes that criterion-referenced measures using spontaneous language samples are increasingly used for testing children's language. This procedure involves, for example, tape-recording a child in conversation or telling a story, and then transcribing and analyzing the data. Looking at individual children, this approach is useful in assisting the tester to evaluate a child's language abilities but is less useful for identifying atypical development since children might avoid the very linguistic constructions that she or he has problems with; using an elicitation testing method could control for that fact. However, comparative studies of these two approaches (elicited vs. spontaneous language samples) suggest that both "deal with the same abilities and that measures derived from [spontaneous] language sample data are more likely to detect progress" (Johnston, 2007, p. 3). McCartney (1993) and Johnston (2006) state that standardized tests are good for identifying problems for subsequent clinical evaluation but that other evaluation procedures provide better guidelines for intervention. For example, naturalistic evaluation can provide better insight into a child's communicative skills (McCartney, 1993). Most clinical evaluations involve standardized and naturalistic testing of expressive language in order to detect problems, followed by making a detailed analysis to describe any problems found, and suggesting areas for intervention (McCartney, 1993).

In the context of methodological issues regarding studies on sign language acquisition, Baker, van den Bogaerde, and Woll (2008) present some relevant points for language testing. They discuss the nature of different language samples, such as spontaneous vs. structured/elicited language. Spontaneous language provides a broad picture of a child's expressive language skills (and also represents more naturally the environment in which the child normally uses language), but it is hard to control for the linguistic behavior under study, that is, the required structures may not occur during a specific video-recording. Elicited language samples are more likely to control the language behavior to be tested, depending on whether a test is aimed at expressive and/or receptive language skill, but the stimuli used must also be considered. For example, using pictures to elicit certain morpho-syntactic structures in a sign language can result in a large variability in the utterances produced, such that it is difficult to be sure of obtaining specific linguistic structures in the data (Haug, 2008b). Tasks that require children to retell a cartoon story seem to provide results that are more consistent and more easily analyzed (e.g., Hoffmeister, 1999).

Haynes and McCallion (1981) raise the question of whether structured language comprehension tests tap only into linguistic understanding or if test results are also affected by other factors, such as motivation or memory. The authors state that it is important when testing language comprehension to be aware of other variables that might possibly affect test performance. Another concern the authors raise is the naturalness of the test stimuli. Testing with single and decontextualized language comprehension items does not resemble language comprehension in a natural environment, and test performance may be affected by memory problems or inattention (Haynes & McCallion, 1981).

Haynes, Purcell, and Haynes (1979) discuss the type of language sampling that is best suited for children aged 4–6 years. They compared the use of conversational vs. picture description tasks in order to investigate which type of task elicits language of greater length and complexity within this age group (Haynes et al., 1979). The results revealed that picture description tasks elicited longer utterances than conversation, but that the conversational condition elicited more complex language use than did the other conditions.

Bates (1993) states that with very young children (up to 24–28 months of age), it is very difficult to test language comprehension using behavioral methods, as young children find the tasks difficult, and as a result, testing may not be reliable. Structured comprehension tests are only considered to be reliable after 28 months (Bates, 1993; Hirsh-Pasek & Golinkoff, 1991; McCartney, 1993).

Another method used in child language acquisition research and in evaluation are parental checklists. The MacArthur-Bates Communicative Development Inventory (Fenson et al., 1993) is designed for use with children aged from 8–30 months. It is widely used and has been adapted in various languages. With these parental checklists, the parents check off which words and early grammatical structures their child produces and comprehends. This methodology shows strong validity and has been confirmed as suitable for use as part of the structured testing of a child's language (Friend & Keplinger, 2008; Johnston, 2007).

It is obvious that deciding upon a certain method depends on various factors, such as the age of the target group, the purpose of the test, and the choice of structured vs. naturalistic approaches. Another important issue is constraints, such as financial resources and the amount of time available for test development or adaptation (McNamara, 2000). These issues are also relevant for the adaptation of the DGS test but the methodological issues have, to a degree, been determined through the BSL Receptive Skills Test.

### 2.1.4 Test Content

Test content is always determined by test purpose. It is essential to determine the test purpose, the test target group and test length, and whether language production and/or comprehension is to be tested. Another important factor is deciding which aspects of a language should be tested (Camarata & Nelson, 2002). Defining what language proficiency is, for example, is very difficult and complex, and also leads to questions of which abilities should and can be tested. Also, it will never be possible to cover all aspects of a language in one test (Bachman, 1990). In order to get a full picture, several tests, each covering different aspects of language, must be included (Wiig & Secord, forthcoming). For tests of language acquisition, testing is mostly concerned with the nature and level of acquisition of linguistic content, form, and use. Content focuses on – among other categories – receptive and expressive vocabulary acquisition and knowledge of abstract concepts. The evaluation of language form is concerned with the acquisition of words and sentence formation rules (morphology and syntax). Tests of language-use (pragmatics) focus on how language is used in different contexts (Wiig & Secord, forthcoming).

In relation to this study, the test content of the BSL test of morphology and syntax specifies the test content of the adapted DGS test.

### 2.1.5 Expressive and Receptive Language Skills

McCartney (1993) describes that the problem with expressive language tests is that they "try to assess an assumed underlying language ability using examples of language performance. Even the most accurate, life-like assessment of a child's expressive language can only give data on which inference of a child's language ability can be made" (p. 36).

In testing language comprehension within a clinical evaluation, it is very important to understand the relationship between comprehension and production, to determine what a child knows about language, and to identify children at risk (Friend & Keplinger, 2008). The challenging task is to test language comprehension (and also to interpret if and how the results contribute to our understanding of language development) since language production constitutes a more easily observable behavior (i.e., it is "just there"). In testing very young children, tests of comprehension may become a measure of compliance rather than of language comprehension (Hirsh-Pasek & Golinkoff, 1996).

There are different reasons why researchers study language comprehension. First of all, it provides a better insight into children's emerging language systems, that is, testing comprehension reveals children's knowledge of a particular language structure before they are actually producing it. Secondly, language comprehension provides an alternative window into the processes of language acquisition; by the time children start to produce a particular structure, they have already acquired it. "Yet the steps leading up to the analysis and mastery of that structure would be less visible without studies of comprehension" (Hirsh-Pasek & Golinkoff, 1996, p. 106). Comprehension is not merely a process that involves mapping sentences to meaning (Brooks, 2004), but rather, it is an interactive process that involves different sources (Hirsh-Pasek & Golinkoff, 1991). Briefly, it is important to look at all the different aspects that might contribute to language comprehension and also to pay close attention to those sources in the development of language comprehension measures which can provide a realistic picture of the child's knowledge of language.

Since the testing of receptive language skills is more difficult – especially in younger ages (up to 28 months of age) – and is therefore a generally neglected area in child language research (Bates, 1993), it is important to look at tests that evaluate sign language comprehension in greater detail.

### 2.1.6 Test Items

Broadly speaking, there are two types of item formats: (1) *selected-response*, and (2) *constructed-response* (Osterlind, 2001). In a selected-response test, the subject is given the correct answer and one or more alternative answers. The alternative answers are labeled *distractors*. The subject has to choose one of the provided answers (e.g., multiple-choice or true-false test items). In contrast, in constructed-response test items, the subject is not provided with any answers and she/he has to produce a word or an entire sentence to answer this test item. An example of this would be short-answer questions where a subject has to write down an answer to indicate reading or listening comprehension (Davies et al., 1999). Test developers should be aware that the response method chosen directly determines which type of item which can be used and may also have an effect on the subject's score (Alderson, Clapham, & Wall, 1995).

The type of item format in the BSL test is a selected-response format with multiple-choice item answers. The same format is used in the adapted DGS test.

### 2.1.7 Pilot Study and Main Study

Before a test can be standardized, it is important to conduct a pilot (Alderson et al., 1995; Fulcher & Davidson, 2007). Based on the results of this pilot, items may need to be revised before a main study can be conducted. One of the main problems in piloting is deciding on how many people a test should be trialed or standardized on. In general, the more subjects, the better. It will not always be possible to trial tests – especially in small populations – with 1,000 children. It is important that the trial be administered as if it were a real test. Issues concerned with piloting and with the main study are equally important for the adapted DGS test, but the issue of sample size is problematic because of the small population of Deaf children. A pilot or standardization study with 1,000 or more Deaf children in any European country will never be possible. This issue will be further addressed later.

Following the piloting and main study stages, and the investigation of validity and reliability, a test is ready to be revised into a standardized format prior to publication.

### 2.1.8 The Rating Method and the Tester

Developing a test also involves deciding on a rating method (which in turn depends on the chosen testing method) and on the training of the tester. For receptive sign language tests, score sheets where a child's answer can be checked off can be used (e.g., Herman et al., 1999; Hoffmeister, 1999; Prinz, Strong, & Kuntze, 1994). For computer-based receptive skills tests, children can click on an answer and the results are stored automatically (Hermans et al., 2010; Mann, 2008).

For language production data, Baker et al. (2008) suggest taking various approaches depending on the linguistic level of a specific structure (e.g., phonology or morpho-syntax) to be analyzed and then choosing an appropriate transcription method (for a detailed discussion, see Baker et al., 2008). Another option is also to develop score sheets, where the tester checks off the occurrence of specific linguistic structures as stated in the test content (Maller et al., 1999). These issues are important for the adapted DGS test, but as it has a computer-based format, the receptive scores are saved automatically.

### 2.1.9 Testing Environment

Bachman (1990) points out different aspects of the testing environment that are important to take into consideration. He discusses the following four facets of the test environment: (1) familiarity of the place and equipment; (2) personnel involved in the test; (3) time of testing; and (4) physical conditions. The test location may be familiar or unfamiliar, and depending on the participant, it may be more or less threatening. Unfamiliar equipment, such as that used in computer-based tests, will have an impact on test performance compared to familiar equipment such as paper-and-pencil tests. Whether test personnel are known or unknown may also have an impact on the results. When the test administrator is familiar, this usually results in better test performance. Results in test performances will differ according to whether tests are conducted by superiors or by peers. The time of testing also plays an important role: Testing in the morning is always better than later in the day (Bachman, 1990). Another factor that can have an impact on the test performance is the physical conditions of the environment, such as noise, humidity, temperature, seating arrangements, and lighting. These issues apply equally to this current study.

Bornstein, Haynes, Painter, and Genevro (2000) conducted a study with two-year-old infants investigating the consequences of variation in person (mother vs. researcher) and place (home vs. laboratory) on children's spontaneous speech. Bornstein et al. (2000) collected data in four settings: (1) mother-child interaction at home; (2) researcher-child interaction at home; (3) mother-child interaction in the laboratory; and (4) researcher-child interaction in the laboratory; thus, they differentiated settings by different degrees of naturalness for the children. Each setting was recorded for later analysis and three different measures of child language were applied. The results of the three language measures showed no differences under naturalistic conditions as compared to the strange condition of a laboratory (Bornstein et al., 2000). However, children spoke consistently more with their mothers than with a stranger/researcher (Bornstein et al., 2000). These results show that it is important for child language researchers to know how linguistic structures and functions may be influenced by contextual and interpersonal factors (Bornstein et al., 2000). Even though the target age group of this present study is older (> 4 years old), these findings are important for the purpose of testing Deaf children with the adapted DGS test.

### 2.1.10 Psychometric Issues in Test Development

Test developers need to provide evidence for the effectiveness of their instruments based on appropriate psychometric measures. While the measures in test construction and development that have been reported in the literature show variation (e.g., Fisseni, 2004; Kline, 2000; Lienert & Raatz, 1998), they all serve the purpose of evaluating a test instrument and/or providing information on the test behavior of participants. The measures most commonly applied to describe how participant behavior relates to the evaluation of his/her performance are reliability, validity, and standardization.

#### 2.1.10.1 Reliability

*Reliability* refers to whether the test actually measures what it is intended to measure (Rust & Golombok, 2000). Reliability can be measured in a number of ways, but two types of evidence are most commonly reported on by researchers. The first is stability over time, and the second is internal consistency. The reliability of a test over time is known as *test-retest reliability* (Kline, 2000). The subjects' scores, which have been obtained on two different occasions, are correlated. The higher the correlation, the more reliable the test is. The *internal consistency* of a test refers to "the degree to which scores on individual items or group of items on a test correlate with one another" (Davies et al., 1999, p. 86). A measure of internal consistency includes statistical procedures such as Cronbach's alpha.

*Inter-rater reliability* refers to the level of agreement between two or more raters on a participant's performance (Davies et al., 1999); for example, by video-recording a child's language production and then comparing the scoring of specific grammatical structures by two different raters. *Intra-rater reliability* refers "to the extent to which a particular rater is consistent in using a proficiency scale" (Davies et al., 1999, p. 91) on different occasions. Intra-rater reliability can be established by comparing the rated scores of candidates that have been tested, for example, on two separate occasions within the span of one month (Davies et al., 1999).

#### 2.1.10.2 Validity

The core claim for the validity of a test is that it really does measure what it claims to measure (Kline, 2000). With regard to Deaf subjects, this could mean whether a test of sign language vocabulary really measures vocabulary knowledge and not, for example, the ability to guess the meaning of iconic signs, which could be equally achieved by non-signing hearing children who have no vocabulary knowledge (e.g., White & Tischler, 1999). There are several validity types: *item* or *content validity*, *concurrent validity*, *predictive validity*, and *construct validity*. Each of these types of validity require different evidence or judgment.

Item or content validity deals with whether, for example, the test items (and the test as a whole) represent the linguistic structures to be tested (Davies et al., 1999). One of the prerequisites for assuring item or content validity in a test of sign language skills is the close collaboration with Deaf native signers during the developmental stage (Singleton & Supalla, 2003). Concurrent validity can be shown by a high correlation between the targeted test and another test that measures the same variable or construct. However, given the very small number of sign language tests, this kind of comparative psychometric measure is difficult to carry out. An example of predictive validity would be a high correlation between the results of a sign language proficiency test and the results of a standardized literacy test, which would indicate that sign language proficiency is a predictor of literacy skills. Construct validity of a language test provides an indication of the extent to which the test instrument represents the theory of language learning that serves as the underlying construct (Davies et al., 1999). The evidence for construct validity "refers to the judgmental and empirical justifications supporting the inference made from test scores" (Chapelle, 1998, p. 50).

Only a few tests for ASL (or other sign languages) have any measures of reliability and validity (Haug, 2008a) compared to the number of tests for spoken English, such as the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2006). This a major difficulty for current research into sign language test development and adaptation.

#### 2.1.10.3 Standardization

An additional issue that can affect the psychometrics of a test is the process of standardization. The success of this process depends on several variables including (1) the size of the population that the sample represents (here, the population of Deaf children), and (2) the homogeneity of the population (Kline, 2000); (e.g., the extent of differences in parent hearing status and diverse linguistic backgrounds).

The key element in successfully determining the quality of a test is the detailed documentation of psychometric properties in test development or adaptation, and as such, it is a crucial element of this current study. Such documentation needs to be presented in a format that facilitates the standardization of the instrument.

### 2.1.11 Use of New Technologies

The use of computer-based testing (CBT) or web-based testing (WBT) in language testing has become increasingly common. One example is the Test of English as a Second Language (TOEFL), where the test is not presented in an examination booklet, but on a computer screen where the candidates have to prompt their responses (McNamara, 2000).

Many paper-and-pencil tests have been converted to computer-based or web-delivered tests. Test content may not change, but two of the main advantages are that scores are stored automatically and that subjects can receive a report of their performance immediately after completing the test. Choi, Kim, and Boo (2003) compared the same language test delivered in paper-and-pencil and computer-based versions and found that scores from both formats were comparable.

The use of CBT and WBT has also lead to new innovations in testing in general. These innovations refer to the content, such as more interactive items with videos, or with the structure of the test, and especially with the scoring and the immediate feedback on the test performance of the participants (Bartram, 2006). Chapelle and Douglas (2006) suggest that the use of multi-media files, such as video and audio, improve the authenticity of the test materials.

Before the Internet began to be widely accessible, CBT was provided at specific locations. WBT raises many issues (such as the effect of technical problems), but one of the main concerns about web-based testing is that of security on a number of levels: (1) security of the test itself (e.g., item content, scoring rules); (2) subject's identity, in relation to authenticating a person's identity and also preserving confidentiality; and (3) security in regard to test results, that is, ensuring that only those who are accredited have access to the data.

An important factor when using CBT or WBT is determining the subject's familiarity with the use of a computer, and this factor needs to be taken into account before beginning test development. The use of a computer, even by young children, has become so widespread that familiarity is less of a concern than a few years ago (Chapelle & Douglas, 2006). These issues are also of importance for this study since the adapted DGS test has a computer-based format that requires the children to know how to use a computer and a mouse.

The use of new technologies for the instruction and testing of both Deaf children and adults is very common today. Examples include web-based video lectures for Deaf students in Slovenia (Debevc & Peljhan, 2004), a computer-based psychiatric diagnostic interview in ASL (Montoya et al., 2004), a computer-based test for Deaf children and young adolescents in DGS (Mann, 2008), a web-based version of the Test of American Sign Language (see Haug, 2008a), and the development of a computer-based environment in ASL for delivering performance-based content from kindergarten through high school (Hooper, Rose, & Miller, 2005; Miller, Hooper, & Rose, 2005). The increasing use of new technologies provides a good opportunity to exploit the use of video to meet the modality-specific features of sign languages in testing. The availability of new technologies has also contributed to the design of the test interface for the adapted DGS test.

### 2.1.12 Diversity in Language Testing

The increasing diversity of the cultural and linguistic backgrounds of children in Europe and in other countries challenges traditional approaches to language testing (Johnston, 2007; Menyuk & Brisk, 2005). Diversity makes it difficult to determine the expected course of language development in bilingual children (Johnston, 2007). As a result, it is important to develop parallel testing instruments for this group of bilingual and multilingual children so that it is possible to measure their development in both their stronger and weaker languages. Another important factor is that cultural influences, attitudes towards testing, and definitions of language proficiency are just a few issues that need to be considered for a fair evaluation of bilingual children's language proficiency (Menyuk & Brisk, 2005). These issues are also of concern for Deaf children. An increasing heterogeneity within Deaf communities has been reported in many countries (e.g., US: Christensen & Delgado, 1993; Gerner de Garcia, 2000; Germany: Große, 2004; Haug & Mann, 2007; Mann, 2008).

In the following section, different models for transferring tests across cultures and languages will be presented.

## 2.2 Models for Transferring Tests Across Cultures and Languages

In research on cross-cultural test adaptation (e.g., van de Vijver & Leung, 1997a), one of the main questions is whether the same instrument can be used with all cultural groups. Is the *construct represented* in both cultures the same way, or are there major differences, or does an overlap exist? Depending on the answer to this question, cross-cultural research suggests three models for transferring a source test to the target culture and language (van de Vijver & Poortinga, 2005): (1) *application* of a test; (2) *adaptation* of a test; and (3) *assembly* of a test.

The application of a test refers to a more or less literal translation of the source language version of the test into the target test version, without any modifications, if "a linguistically appropriate translation turns out to be psychologically adequate" (van de Vijver & Poortinga, 2005, p. 52). This assumes that the construct measured is fully equivalent in both cultures. Adaptation refers to the case when the construct is not fully covered in the target culture, and the test can only be adapted by "rephrasing, adding, or replacing items that measure the missing aspects" (van de Vijver & Leung, 1997a, p. 265). Adaptation should be applied when the construct is not entirely represented in the target culture. But the changes caused by an adaptation require that the underlying construct be measured equally in the source and the target test (van de Vijver & Leung, 1997b). An example of the use of adaptation is the Minnesota Multiphasic Personality Inventory (MMPI) which has been adapted to Chinese (Cheung, 1985). Some of the original American items were meaningless in the Chinese context and were therefore modified to the Chinese context. But most of the original MMPI items were retained. Those were items relating to social interaction, activity level, nonchalant attitudes, modesty, sex, and admission of personal problems (Cheung, 1985). Cheung (1985) interprets these differences as cultural differences in social norms and values. Assembly refers to situations where it is necessary to assemble a new test because the construct representation in the original instrument is inadequate for the target culture and as a result, "a new instrument is developed to capture the construct more adequately in the new cultural context" (van de Vijver & Leung, 1997b, p. 36). In this case, the newly assembled test should still cover the same underlying construct. One of the few examples of this model is a study by Church (1987), where he adapted a Western personality test which was unable to capture many of the indigenous personality constructs of Filipino culture.

Looking at these different approaches (application, adaptation, assembly; van de Vijver & Poortinga, 2005), it is suggested that the approach of adaptation is the most appropriate for the present study: to transfer a test from a source to a target sign language and to take into consideration the target language-specific structures and specific concepts of the target culture in the adaptation process while still measuring the same underlying construct (i.e., language development). The first approach listed above (application), which involves translation from the source to the target language, will not be proposed here. There is evidence from the translation of spoken language tests that this is not a promising approach (Alant & Beukes, 1986; Chavez, 1982; Rosenbluth, 1976; Simon & Joinier, 1976). The third approach (assembly) will also not be proposed here as the construct (representation) under investigation is not sufficiently different to warrant a new assembly. It is therefore argued that adaptation constitutes the most promising approach for "transferring" a test across sign languages. Adaptation has also been found to be a successful approach for spoken language tests (Friend & Keplinger, 2008; Hamilton, Plunkett, & Schafer, 2000; Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993; Maital, Dromi, Sagi, & Bornstein, 2000; Thordardottir & Ellis Weismer, 1996).

## 2.3 Test Adaptation

Having argued for test adaptation as the most promising approach for transferring a test from a source sign language to a target sign language, it is important to define the term *adaptation* in order to clearly distinguish it from the term *translation*. Hambleton (1994, 2005) and Hambleton and Patsula (1998) define adaptation as the entire process beginning with the source test and ending with the target test, whereas translation is only one step within this process (i.e., to translate test instructions or individual items into the target language). Other researchers also emphasize the need to distinguish the two terms. Geisinger (1994) uses the term *adaptation* rather than *translation* when referring to the transfer of a test from one language to another. Adaptation takes into account both linguistic and cultural differences and involves flexibility in test construction.

Oakland & Lane (2004) illustrate the many factors inherent in the adaptation process:

> Test adaptation refers to a process of altering a test originally designed for use in one country in ways that make the test useful in another country. The immediate goal in adapting the test is to develop a parallel test (i.e., target test) that acknowledges the linguistic, cultural, and social conditions of those who will be taking the adapted test while retaining the measurement of the constructs found in the original (i.e., source) test. The ultimate goal is to have two tests that measure the same trait in fair, equitable, and somewhat equivalent fashion. (Oakland & Lane, 2004, p. 239)

Having provided a definition of the terms *adaptation* and *translation*, a closer look will be taken at the adaptations of spoken and sign language tests.

### 2.3.1 Adaptation of Spoken Language Tests

There is one major difference between psychological tests and language tests: In psychological tests, language constitutes the vehicle of communication between the test and the subjects taking the test, that is, language rarely constitutes the trait that is being tested (Oakland & Lane, 2004), whereas in language tests, the trait being tested and the vehicle of communication between the test and the subject are identical.

Compared to the amount of literature on the adaptation of psychological tests, very few studies are reported on the adaptation of language tests.

Studies on spoken language test adaptation and translation can be summarized into two broad approaches. The first is a translation of an existing source test into a target language; the second – and generally more successful approach – is adaptation. The criterion for success in this context is whether children perform in a comparable way on the original and translated versions.

In most of the studies that have used a translation approach (e.g., Alant & Beukes, 1986; Chavez, 1982; Rosenbluth, 1976; Simon & Joinier, 1976), children perform more poorly on the translated test version than on the original and there is lower reliability. Possible explanations are that in the translation process, (1) culturally sensitive or relevant concepts that play an important role in the lives of children in the target culture are not attended to, (2) the path of language development in one language – as it is represented in the source test – may differ in other languages (Slobin, 1973), and (3) the experiences of children vary across cultures (Thordardottir & Ellis Weismer, 1996).

Studies that choose an adaptation approach (e.g., Hamilton et al., 2000; Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir & Ellis Weismer, 1996) have been far more successful than translated tests. Where cultural and linguistic differences between the source and target languages are addressed (Hamilton et al., 2000; Jackson-Maldonado et al., 1993; Caselli, Casadio, & Bates, 1999), similar patterns of language development can be seen across languages.

In adaptations, test developers use (1) data from language acquisition studies and studies of specific features of the language (e.g., Friend & Keplinger, 2008; Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir & Ellis Weismer, 1996), (2) omit and add items in order to cover cultural relevant terms and concepts (e.g., Hamilton et al., 2000; Jackson-Maldonado et al., 1993), and (3) include external input, such as a panel of language experts (e.g., Maital et al., 2000; Pakendorf & Alant, 1997) to review the adapted test version and/or feedback from parents after piloting. It is important to note that most of the adaptations mentioned here have been based on the MacArthur-Bates CDI (Fenson et al., 1993).

Several of these studies (e.g., Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir & Ellis Weismer, 1996) report that the most significant changes required were in the second part of the CDI (the sentence part), where differences among languages in early syntactic development are more clear-cut than in the first part (early vocabulary). The complexity of a linguistic feature represented in a test – ranging from vocabulary to complex morphological structures – affects the extent to which adaptations are required: Vocabulary items are often easier to adapt than complex morphological, language-specific structures. Other issues that are relevant for the adaptation and later use of language tests relate to culture-determined experiences and the resulting behavior of children in testing situations: (1) for example, the use of computers for testing (Friend & Keplinger, 2008); (2) the culture-appropriate learned behavior of a child towards an adult in a testing situation (Pakendorf & Alant, 1997); (3) the use of a tester from the same cultural group as the children to be tested (e.g., Norris, Juárez, & Perkins, 1989; Pakendorf & Alant, 1997; Solarsh & Alant, 2006); and (4) the need to attend to non-mainstream variants or dialects of the target language (e.g., Alant & Beukes, 1986; Norris et al., 1989; Restrepo & Silverman, 2001).

Mueller Gathercole and her colleagues (2008) also address the need, when developing tests for bilingual populations, to ensure that norms take the different language experiences of the children into account.

Compared to the option of developing a new test, the adaptation of language tests is attractive. For cross-linguistic research, where the goal is to enable comparisons between two language groups, it is important to look at the extent to which such comparisons are possible. Depending on the closeness of the source and target test versions, comparisons may be more or less possible on different levels. In relation to adaptations of the CDI, for example, it is not possible to compare the performance of the children on normed scales.

Language test adaptation is a more successful approach than simple translation, but it is also clear that in terms of comparability, the goal of most adaptations is to test a general ability (e.g., receptive vocabulary), rather than to compare the scores of tests in two different languages (e.g., Friend & Keplinger, 2008).

### 2.3.2 Adaptation of Sign Language Tests

The situation for the adaptation of sign language tests is quite different from that for spoken languages. The current state of research for many sign languages does not provide the same possibilities for adaptation as discussed above in relation to the adaptation of spoken language tests. Only one publication directly addresses the issue of adaptation from a source to a target sign language test (Haug & Mann, 2008). One of the key issues is concerned with the psychometric properties that need to be established in an adapted test, even when the source test shows strong evidence of reliability and validity (Hambleton, 1994, 2001, 2005).

Potential problems in the adaptation of a test from one sign language to another can be summarized into two broad categories: (1) *language-specific issues*; and (2) *culture-related issues*. In relation to language-specific issues in the adaptation of the Test Battery for ASL Morphology and Syntax (Supalla et al., 1995, unpublished) to Auslan (Schembri et al., 2002), morpho-syntactic differences were found. For example, derivationally related noun-verb pairs showed greater variability in Auslan than in ASL. A similar observation was made by Johnston (2004) in adapting the BSL Receptive Skills Test to Auslan (although the two languages are closely related). The BSL signs WRITE and PENCIL showed a derivationally related noun-verb distinction while in Auslan, the signs for these two referents were derivationally unrelated. Given that there are only 40 test items, this might make the pilot Auslan test easier than the BSL test it is based on. A similar issue has been reported in relation to the adaptation of the BSL Receptive Skills Test to Danish Sign Language (Haug & Mann, 2008).

Surian and Tedoldi (2005) experienced difficulties in the adaptation of the BSL Receptive Skills Test to Italian Sign Language (LIS), related to morphology and syntax, particularly when trying to adapt structures that involved negation. These difficulties may have stemmed from the wider variety of devices that signers of LIS have at their disposal to express this grammatical feature in comparison to users of BSL. The opposite findings were reported in Haug and Mann (2008) in relation to the adaptation of the same test to French Sign Language (LSF). In this study, the researchers faced the challenge of working with a smaller number of forms of negation in the target sign language, LSF, than in BSL. While the BSL test consists of 40 items, of which 8 represent different forms of negation (e.g., BSL signs such as NOTHING, NO, NOT, NOT-LIKE), LSF has fewer signs to express negation. The effect this had on the adapted version for LSF was item redundancy, as some items ended up measuring the same forms of negation more than once.

As for culture-related issues, they can often be handled by altering stimulus materials to better fit artifacts in the target culture, such as changing pictures of the round red British mailbox to the appropriate image for the target culture (e.g., for Danish Sign Language; Haug & Mann, 2008). Prinz, Niederberger, Gargani, and Mann (2005) compared selected items from two of the six subtests of the Test of American Sign Language (TASL; Prinz et al., 1994) with their adapted versions in Swiss French Sign Language (Niederberger, 2004). Prinz and his colleagues (2005) report results in relation to the participants' responses for one of the items from the story comprehension task concerned with obtaining a driver's license. While most American subjects showed no difficulties with this item, it was reported to be one of the harder items for Swiss French subjects. The researchers hypothesized that this divergence may be because of the different significance of having a car in the two cultures.

Problems during the adaptation process can arise from (1) linguistic differences between the source and the target language, and (2) differences in the source and the target cultures. Both are important aspects that need to be considered in the adaptation of the BSL test to DGS.

## 2.4 Review of Sign Language Tests

Sign language tests can be categorized according to three different objectives (Haug, 2008a): (1) instruments for linguistic research; (2) instruments for educational purposes; and (3) instruments for evaluating sign language acquisition of Deaf children.

### 2.4.1 Instruments for Linguistic Research

Tests in this category are the Test Battery for ASL Morphology and Syntax (Supalla et al., 1995, unpublished); the Test Battery for Australian Sign Language Morphology and Syntax (Schembri et al., 2002), which is an adapted version of the Supalla et al. ASL test battery; the Grammatical Judgment Test of ASL (Boudreault, 1999; Boudreault & Mayberry, 2000); the American Sign Language-Sentence Reproduction Test (ASL-SRT; Hauser, Supalla, & Bavelier, 2006); and the Non-Sign Repetition Task for BSL (Mann, Marshall, Mason, & Morgan, 2010). The first two were primarily designed to obtain information about how specific morpho-syntactic structures are produced by Deaf adult signers. The Grammatical Judgment Test of ASL was developed in the context of a research project investigating the effect of the age of acquisition of ASL on grammatical processing. The ASL-SRT focuses on a verbatim recall of ASL sentences of different complexity in order to differentiate between Deaf subjects with different levels of ASL mastery. This is an adaptation of the Speaking Grammar Subtest of the Test of Adolescent and Adult Language, 3rd Edition (Hammill, Brown, Larson, & Wiederholt, 1994). The Non-Sign Repetition Task of BSL was used in a research project examining the phonological abilities of Deaf and hearing children and Deaf adults. The studies conducted with these tests had a different purpose than the purpose of this current study, which is to adapt a test for sign language development.

### 2.4.2 Instruments for Educational Purposes

Tests in this category include the American Sign Language Assessment Instrument (ASLAI; Hoffmeister, 1999, 2000); the Test of American Sign Language (TASL; Prinz et al., 1994; Strong & Prinz, 1997, 2000) the Computer Test of German Sign Language (CTDGS; Mann 2008); and the adapted version of the TASL to Swiss French Sign Language (Niederberger, 2004, 2008). These tests were designed and used in studies to investigate the relationship between Deaf children's knowledge of a sign language and their literacy performance. They focus on a different purpose than the test of sign language development in this present study.

### 2.4.3 Instruments for Evaluating Sign Language Acquisition

In this category are tests that were developed with the goal of evaluating sign language development in Deaf children across different time spans in different sign languages. These tests cover children within various age ranges from 8 months to 15 years: for example, 8–36 months of age in the adaptation of the MacArthur-Bates CDI (Fenson et al., 1993) to ASL (Anderson & Reilly, 2002), BSL (Woolfe, Herman, Roy, & Woll, 2010), and Sign Language of the Netherlands (NGT; Hoiting, 2009); 2–5 years of age in the Developmental Assessment Checklist of Sign Language of the Netherlands (NGT-OP; Baker & Jansma, 2005); and 3–11 years in the BSL Receptive Skills Test (Herman et al., 1999). Other examples are the Australian Sign Language Receptive Skills Test (PARST; Johnston, 2004), the Signed Language Development Checklist (Mounty, 1993, 1994), the American Sign Language Proficiency Assessment (ASL-PA; Maller et al., 1999), Assessment for Sign Language of the Netherlands (Jansma, Knoors, & Baker, 1997), Assessment Instrument for Sign Language of the Netherlands (Hermans et al., 2010), BSL Narrative Skills Test (Herman et al., 2004), Vocabulary Test for German Sign Language and Written and Spoken German (Perlesko; Bizer & Karl, 2002), and the Aachen Test for Basic German Sign Language Competence (ATG; Fehrmann et al., 1995a, 1995b; Huber, Sieprath, & Werth, 2000). The ATG can also be used for adults.

These instruments focus on different pre-linguistic and linguistic levels: early gestures (Woolfe et al., 2010); phonology (e.g., Hermans et al., 2010; Mounty, 1993); vocabulary (e.g., Anderson & Reilly, 2002; Bizer & Karl, 2002; Hoiting, 2009); morphological and syntactic structures (Fehrmann et al., 1995a, 1995b; Herman et al., 1999; Hermans et al., 2010); and narrative production (Herman et al., 2004). Most instruments test only very specific structures, such as morphological and syntactic structures in the BSL Receptive Skills Test (Herman et al., 1999), but some tests also focus on a variety of structures (e.g., phonology, morphology, and syntax), such as the Assessment Instrument for Sign Language of the Netherlands (Hermans et al., 2010). Some tests focus only on language production (Herman et al., 2004), some only on comprehension (Herman et al., 1999), and some on both production and comprehension (Fehrmann et al., 1995a, 1995b; Hermans et al., 2010).

Interestingly, only very few tests are actually commercially available. That very few tests have been published shows that most of them are still under development. Only the BSL Receptive Skills Test (Herman et al., 1999), the BSL Narrative Production Test (Herman et al., 2004), the Assessment Instrument for Sign Language of the Netherlands (Hermans et al., 2010) and the Perlesko for DGS (Bizer & Karl, 2002) are available. This also reflects one of the weaknesses among sign language tests, as pointed out by Haug (2008a), that is, the lack of reported psychometric properties. To investigate and report the psychometric properties of the adapted test is a very important issue for the adaptation of the BSL test to DGS.

This short overview of the available sign language tests served as a basis for the decision which test would be used for adaptation to DGS. Three important criteria needed to be met: (1) reported psychometric properties; (2) testing of the development of language comprehension, an important but often neglected area in language testing; and (3) focusing on an age group where standardized testing formats can be used (> 3 years). Only the BSL Receptive Skills test met these criteria.

### 2.4.4 Tests for German Sign Language

In Germany, only three sign language tests are available: the ATG (Fehrmann et al., 1995a, 1995b; Huber et al., 2000); the CTDGS (Mann, 2008); and the Perlesko (Bizer & Karl, 2002). The Perlesko is a norm-referenced vocabulary test that evaluates the vocabulary comprehension of DGS, spoken, and written German in elementary school-aged Deaf children ($3^{rd}$ to $5^{th}$ grades), and thus also taps into a different domain of language and a different age group than the intended test adaptation. The CTDGS is a computer-based receptive skills test of referential distinctions in DGS that addresses a different age group (> 8 years) and also has a different purpose than the intended adaptation of the DGS test. The CTDGS has not yet been normed.

The ATG is a criterion-referenced test that can be used with children and adults. The adult version takes four hours to administer, the children's version two hours. It aims to measure basic competence in DGS, defined as "the language competence which an adult deaf native signer would consider to be the minimum level of fluency/knowledge required to be considered as a fluent DGS user" (Haug, 2008a, p. 66). The ATG can be applied to a variety of purposes: (1) diagnosis of language development in Deaf children; (2) monitoring of sign language development in school; and (3) linguistic self-evaluation of Deaf adults. The test can be used for children aged 6 years or older and provides an in-depth investigation of specific linguistic structures, consisting of nine sub-tests that evaluate both expressive and receptive skills, focusing on different linguistic units, such as signs, phrases, and text (Haug, 2008a). The ATG has not been published, and, furthermore, no psychometric properties have been reported. Additionally, the ATG focuses on a different age group (> 6 years) than the intended age group of the DGS test adaptation. It is also a criterion-referenced test, whereas the goal of the DGS adaptation is a norm-referenced test. Also, the ATG is too long to be used in schools.

In sum, the three tests that are available for DGS differ from the planned adaptation of the BSL Receptive Skills Test (Herman et al., 1999) with respect to: (1) target age group; (2) purpose, that is, comprehension of morpho-syntactic structures in DGS; and (3) standardization and ease of use in educational contexts.

In the following section, the template for the test adaptation will be presented.

### 2.4.5 The British Sign Language Receptive Skills Test

The BSL Receptive Skills Test (Herman et al., 1999) is designed for children aged 3 to 11 years. Following a pilot study on 41 Deaf and hearing children between 3 and 11 years (28 children with at least one Deaf parent, and 13 hearing children with a native signing background), the test was revised and standardized on 138 children. The participants in the standardization study included (1) Deaf children with Deaf parents, (2) hearing children of Deaf parents (with a native signing background), and (3) selected Deaf children of hearing parents (identified by teachers) who were enrolled in a bilingual program, had hearing parents with unusually good signing skills, or who had older Deaf siblings. The BSL Receptive Skills Test focuses on selected aspects of morphology and syntax of BSL. It consists of a vocabulary check and a video-based receptive skills test.

*Vocabulary check*: The children confirm their knowledge of the 22-item vocabulary used in the main test through a simple picture-naming task that identifies signs taken from the receptive skills test.

*Receptive skills test:* The video-based Receptive Skills Test consists of 40 items, which are ordered by level of difficulty. Because of regional variation in signs, there are two versions of this task, one for the North and one for the South of the UK. The items of this test evaluate children's receptive knowledge of a variety of BSL syntactic and morphological structures: (1) spatial verb morphology; (2) number and distribution; (3) negation; (4) size/ shape specifiers; (5) noun-verb distinction; and (6) handling classifiers. Table 2.1 gives examples of items from the BSL Receptive Skills Test. A detailed description will be provided later when comparing these structures with DGS.

Studies on the acquisition of BSL and ASL were reviewed in order to identify linguistic features that are important for the acquisition of BSL. The following linguistic structures were identified after the review and are included for item development (Herman, 2002).

Table 2.1: Examples of Items from the BSL Receptive Skills Test

| *Linguistic category* | *Item** |
|---|---|
| Spatial verb morphology | BOX UNDER BED (Item 17) |
| Number and distribution | QUEUE (Item 24) |
| Negation | ICE-CREAM NOTHING (Item 3) |
| Size and shape specifiers (SASS) | CURLY-HAIR (Item 16) |
| Handling classifiers | EAT-THIN-SANDWICH (Item 37) |

* For a complete list of items, see Appendix D-3

*Spatial verb morphology*[5] refers to complex verbs in BSL (Herman, 2002; Herman et al., 1999; Sutton-Spence & Woll, 1999). Space can be used for different purposes in BSL (Sutton-Spence & Woll, 1999): topographic space and syntactic space. In the category of spatial verb morphology, three types of verbs are included: spatial verbs with whole entity classifiers; agreement verbs; and complex AB verb constructions, a subclass of agreement verbs.

Whole entity classifiers make use of topographic space, and the spatial information conveyed is a representation of a referent's actions and loca-

[5] For the sake of clarity, the terminology used throughout this book in relation to the adaptation of the test to DGS is that used in the introductory book on BSL linguistics (Sutton-Spence & Woll, 1999), which is also the terminology adopted for the BSL test.

tions in the real world. In whole entity classifiers, the sign movement starts at the initial location of an object and finishes at its final location (e.g., TAKE-BOOK-FROM-SHELF). The handshape of the whole entity classifier changes depending on the class to which the object belongs (e.g., *vehicle* or *small rectangular object*). Whole entity classifiers can also inflect for manner and aspect[6].

Syntactic space, on the other hand, uses space to convey grammatical information without real world mapping. For example, in the sentence *John gives the book to Mary* in BSL, a location in space is established for *John* and another for *Mary* through indexing, and then the verb representing *who gives what to whom* is signed by performing a movement that starts at the location of *John* and moves to the location of *Mary*. Agreement verbs make use of syntactic space.

In summary, whole entity classifiers, agreement verbs, and complex AB verb constructions, are represented in the category *spatial verb morphology* in the BSL test.

*Number and distribution* are the equivalent of plurals in English, but are morphologically more complex in BSL. To perform number/distribution in BSL, at first the lexical sign representing a referent is produced, followed by a classifier handshape representing the class to which the referent belongs, and a repeated movement. For example, to produce *beds*, at first the lexical sign BED needs to be produced, followed by the Flat-B handshape with a small downward movement repeated as the hand moves from left to right. Thus spatial information is expressed about the location in space of the referents, constituting an overlap with the category of spatial verb morphology (i.e., some items belong to two categories). Plurals can also be expressed by using a number (e.g., THREE), quantifier (e.g., MANY) (Herman, 2002), or by adding a bound plural morpheme (e.g., HOUSE vs. HOUSE++) (Sutton-Spence & Woll, 1999).

*Negation* is performed in BSL with a variety of linguistic devices. These include a negation facial expression, head turns and headshakes, specific negation signs (such as NOTHING, NOT, NEVER), and changes in how a sign is articulated (e.g., the addition of negative affixes to verbs such as LIKE or KNOW) (Sutton-Spence & Woll, 1999). Specific negation signs and

[6] Two of the items (Items 18 and 29) in the category of *spatial verb morphology* can be considered as agreement verbs, since syntactic information about subject and object is conveyed by the start and end locations of the sign.

signs such as NOT-LIKE must be used in combination with the negation facial expression and/or head turns.

*Size and shape specifiers (SASS)*: these are a subtype of classifier in BSL that identify the size and shape characteristics of nouns. SASSes are often used in contexts where in English adjectives would be used to modify a noun (e.g., *small square spots*) (Herman et al., 1999).

*Noun/verb distinctions*: a group of nouns and verbs in BSL are derivationally related, such as AEROPLANE / FLY and CAR / DRIVE. In most pairs, the noun has a short movement, which ends with the sign being held briefly, while the verb "has a longer movement which tapers off" (Herman et al., 1999, p. 5).

In *handling classifiers,* the handshape represents how an object is held. They can be found in predicates (e.g., in sentences like *the boy ate a pizza*, *the boy ate a hamburger*, *the boy ate chips*, etc.). The handshape in *ate* varies according to how the object is usually handled.

#### 2.4.5.1 Testing Procedure of the BSL Receptive Skills Test

The BSL Receptive Skills Test is presented to participants in video format. In addition to the test items, it also includes signed instructions. This format facilitates a standardized presentation of the test and reduces demands on the tester. The vocabulary check, however, is administered live and requires some BSL skills on the part of the tester.

#### 2.4.5.2 Psychometrics of the BSL Receptive Skills Test

In order to establish test-retest reliability for the receptive task, 10% of the sample on which the test was standardized were retested. The test scores improved on the second testing, but the rank order of scores was preserved. There was also a high correlation (.87) between the test and retest scores. Split-half reliability analysis for the internal consistency of the receptive test revealed a high correlation (.90) and, therefore, represents a high internal consistency. The scores for the BSL Receptive Skills Test of the children involved in the pilot were compared with those of subjects not previously exposed to the test materials. There was a slight advantage in the pilot children, however, the difference between the groups did not achieve statistical significance ($p = .70$).

The BSL Receptive Skills Test has been adapted to Australian Sign Language (Johnston, 2004), Danish and French Sign Language (Haug & Mann, 2008), Italian Sign Language (Surian & Tedoldi, 2005), American Sign Language (Enns & Zimmer, 2009), and Maltese Sign Language (Haug, 2008a). Adaptation to Japanese Sign Language is currently underway (R. Herman, personal communication, April 28, 2009).

## 2.5 Sign Language Acquisition

In this section, studies on sign language acquisition will be analyzed. Since the age group of the adapted DGS test is from 4 years onward, a short overview of sign language acquisition from birth to 4 years old will be presented. This is followed by an extended literature review on the acquisition of the linguistic structures represented in the BSL test (e.g., negation, number and distribution). A review of studies of DGS structures that should be represented in the DGS test, and selected studies on cross-linguistic differences in order to evaluate the suitability of the test for adaptation, precede the review of the acquisition studies.

Language acquisition proceeds through different stages. Children have the potential to acquire any language they have access to. Based on research on sign language acquisition, it can be assumed that Deaf and hearing children pass through similar paths of development (Woll, 1998).

Research on sign language acquisition from birth to 4 years old is summarized in the Table 2.2, while Table 2.3 provides an overview of the acquisition of spoken English from 0–7 years.

For the development or adaptation of a test that tests a Deaf child's sign language development relative to his/her peers, it is important to look at language acquisition studies. Because of the paucity of acquisition studies of DGS (only one study is available: Hänel, 2003), an overview of other, better-documented sign languages (e.g., ASL, BSL) will be provided first, followed by an overview of DGS acquisition research.

Table 2.2: Overview of Sign Language Acquisition (Approx. 0–4 Years Old)

| *Age span* | *Structure(s)* | *Description* | *Author(s) & year (examples)* |
|---|---|---|---|
| 0–12 months | Babbling & gestures | After a few months of vocal babbling, vocal babbling decreases and Deaf children of Deaf parents start to babble manually. | Masataka (2000); Petitto & Marentette (1991) |
| 0;9–2;3 yrs | Pointing & pronominal reference | Deaf children use non-linguistic pointing to indicate present people, objects, and location from 9–12 months. They stop pointing referring to people between 12 and 18 months of age, but they continue to use pointing referring to objects and locations. The pointing to people returns at about 18 months of age, but now as pronouns. But full control of personal pronouns YOU and ME is not achieved until the age of 25–27 months. | Hatzopoulou (2008); Petitto (1987) |
| 0;9–3;0 yrs | First signs | The productive vocabulary development of Deaf children: about 4–8 signs between 8–11 months; up to 60 signs between 12–19 months; between 150–200 signs between 20–27 months; and a constant vocabulary growth with about 300–380 signs by age 30–36 months. The first signs are produced in isolation and in uninflected citation form. | Anderson & Reilly (2002); Hoiting (2009); Woolfe et al. (2010) |
| 1;6–1;11 yrs | Early morphology & syntax | The first verbs appear in the lexicon, but no productive verb morphology used (i.e., verbs appear only in citation form, that is with no subject or object agreement in agreement verbs, and no use of classifiers in spatial verbs). The first two–sign utterances appear. But the use of sign space is still absent. | Morgan, Barrière, & Woll (2003); Maller et al. (1999) |
| 2;0–2;5 yrs | Phonology, pronominal reference, & morphology | Between 2;0 and 2;5, the phonology still differs greatly from that of adult signers. There seems to be universal pattern in the acquisition of handshape development, with unmarked handshapes, such as pointing hand, flat hand, and fist, appearing first. The use of pronominal reference is further extended in the age span from | Boyes Braem (1994); Meier (1987); Newport & Meier (1985); Petitto (1987) |

| *Age span* | *Structure(s)* | *Description* | *Author(s) & year (examples)* |
|---|---|---|---|
| | | 2;0 to 2;5. Children use pointing to an addressee (YOU) at about two years of age. Some children show evidence for reversal errors (i.e., they sign YOU, but referring to oneself). By 2;5 the pointing to first, second, and third person is used correctly. Verb agreement is being used, but mostly in citation form (with a short movement in space), with omitted agreement and not picking a particular referent. | |
| 2;6–2;11 yrs | Morphology | Parts of the morphological subsystem of classifiers used in spatial verbs appear, but rather as unanalyzed wholes, without productive use. The first productive use of verb agreement starts in this age span. Most of morphology has been acquired by age 2;6 and 3;0. | Morgan et al., (2003); Morgan, Barrière, & Woll (2006); Newport & Meier (1985) |
| 3;0–3;5 yrs | Morphology | The use of inflectional morphology of spatial verbs for movement and manner occurs, but they are not yet combined, and either the manner or the movement morpheme is omitted. Verb agreement is mastered, but only where reference is made to present objects in the environment/real–world locations. The omission of verb agreement with abstract spatial loci (i.e., with nonpresent object, continues till after 3;0). The first correct use of some number and aspect morphemes are found with spatial and agreement verbs. | Bellugi, van Hoek, Lillo-Martin, & O'Grady (1988); Meier (2002); Morgan et al. (2003); Morgan & Woll (2002b) |
| 3;6–3;11 yrs | Phonology & morphology | Between 3;6 to 3;11, children use lexical compounds, but they are not produced with the characteristic phonological pattern. Both spatial and agreement verbs now have both manner and movement, but they are produced sequentially rather than simultaneously. However, by 3;11 they start to coordinate the usage of both. Children have not yet acquired the establishment of abstract loci. | Newport & Meier (1985); Woll (1998) |

Expanded version of Woll (1998)

Table 2.3: Overview of English Language Acquisition (0–7 Years Old)

| *Age: 0–6 months* | *Age: 7–12 months* | *Age: 1;1–1;6 yrs* | *Age: 1;7– 2;0 yrs* | *Age: 2;1–2;11 yrs* |
|---|---|---|---|---|
| - Babbling begins<br>- Babbles to self, others, and objects<br>- Vocalizes pleasure and displeasure<br>- Seeks sound source<br>- Occasionally vocalizes in response to speech | - Babbles in repeated syllables, such as bababa<br>- Uses /m, n, t, d, b, p, z/ when babbling syllables<br>- Understands parental gestures<br>- Looks at common objects and familiar people when named<br>- Listens to and imitates more sounds<br>- Uses most sounds in (C&V) in vocal play, beginning of phonetic drift<br>- Shows evidence of first true word (10-18 months) (1-3 words)<br>- Responds to own name<br>- Shakes head for no<br>- Waves bye bye | - Combines gestures and vocalization<br>- Uses between 3-20 words<br>- Uses most sounds, but speech continues to be unintelligible<br>- Follows simple one part directions using in/on<br>- Points to named objects in real life or pictures<br>- Points to some body parts<br>- Uses mostly nouns to speak<br>- Interacts for requesting objects or attention | - Uses more words than jargon<br>- Uses intonation to denote a question<br>- Understands 200-300 words<br>- Speaks about 50 words<br>- Names some animals with associated sounds<br>- Refers to himself/herself by name<br>- Starts to combine nouns and verbs<br>- Combines 2 words, such as mummy car, big bus etc<br>- Displays more turn taking with communicative partners<br>- Understands yes-no questions<br>- Answers to what's that? questions | - There is an increase in utterance length of 3-4 words (I want to play, daddy drive car)<br>- Child uses approximately 200-500 words<br>- Child understands up to 900 words<br>- Response to some yes-no questions<br>- Asks 1- to 2-word questions<br>- Understands some wh-questions (why, what, where)<br>- Grammatical structures start to emerge, such as articles a and the; prepositions in and on; plurals cats, boats, cars; regular past tense walked; negative not<br>- Begins to use pronouns (I, me, he, she) |

| *Age: 3;0–3;11 yrs* | *Age: 4;0–4;11 yrs* | *Age: 5;0–5;11 yrs* | *Age: 6;0–7;0 yrs* |
|---|---|---|---|
| - Understands 1,200-2,000 words<br>- Uses 800-1,500 words<br>- Utterances are generally 4-5 words long<br>- Knows *in*, *on*, *under*, *big*, *little*, matches colors<br>- Begins to count<br>- Sentence grammar improves although some errors persist<br>- Uses */s, z/* as plural markers (*cats*, *chairs*)<br>- Understands comparison<br>- Questions stage: asks and answers simple questions (*who*, *what*, *where*, *why*)<br>- Tells two events in chronological order<br>- Consistently uses regular plurals, possessives, and simple past tense verbs | - Understands up to 2,500 words<br>- Uses 1,500-2,000 words<br>- Should have speech that is mostly understood by others<br>- Masters the */f, k, g/* sounds<br>- Masters many blends (*/st, sp, sm, sn/*, etc)<br>- Continues to understand spatial concepts, e.g. *top, bottom, front, back*<br>- Listens to short simple stories<br>- Answers complex 2-part questions<br>- Uses 5-8 words per utterance<br>- Counts to ten<br>- Can discuss emotions and feelings<br>- Uses more helping verbs like *could, would*<br>- Using more word endings (*bigger, helped*).<br>-Tells stories as a sequence of events | - Uses the sounds */l, r/*<br>- Understands 12,000 words or more<br>- Understands opposites<br>- Understands more time and quantity concepts<br>- Uses all pronouns correctly<br>- Uses proper sentence structures most of the time<br>- Begins using adverbial endings (*slowly, faster*)<br>- Uses utterances that average 6 words<br>- Tells stories with a main character and sequence of events<br>- Uses conjunctions (*and, but, or*)<br>- Uses past tense (climbed) and future tense appropriately | - Uses */th, sh, ch, j/* sounds<br>- Understands approximately 25,000 words<br>- Counts to 100<br>- Uses most morphological marker consistently<br>- Can use irregular forms of words (*rode*, *drew*, *mice*) more consistently<br>- Speaks with an average utterance length of 7 words.<br>- Tells stories with main character, plot and sequenced events leading to an ending<br>- Uses passive voice appropriately |

From: Gard, Gilman, & Gorman (1993)

In the previous section, an overview of sign language acquisition from 0–4 years old was presented, mostly based on reviews of acquisition studies for ASL and BSL.

### 2.5.1 Comparison of Linguistic Structures: Cross-Linguistic Differences and Sign Language Acquisition

In the following sections, studies on the structures represented in the adapted DGS test will be presented, followed by a comparison with cross-linguistic research to point out similarities and differences across sign languages. Following the discussion of each structure across sign languages, studies on sign language acquisition addressing the linguistic structures that are represented in the adapted DGS test (e.g., negation, handling classifiers) will be presented. The reviewed and analyzed acquisition studies are mostly based on ASL and BSL. The following categorization of linguistic structures is used: (1) spatial verb morphology, including agreement verbs, complex AB verb constructions, and spatial verbs with whole entity classifiers; followed by (2) SASSes; (3) handling classifiers; (4) number and distribution; and (5) negation. Whole entity classifiers, SASSes, and handling classifiers will be presented together in one section.

Noun-verb derivational morphology – as represented in the BSL test – will not be presented here since derivationally related noun-verb distinctions in DGS appear not to exist in DGS (Becker, 2003). However, in order to confirm or disconfirm these claims, the noun-verb derivational morphology items were also adapted to DGS for the pilot.

Table 2.4: Overview of Selected Studies of Sign Language Acquisition

| *Linguistic Structures* | *Language* | *Emergence and / or mastery of structure* | *Methodological issues* | | | | | *Cross-linguistic comparison* | *Author(s) & year* |
|---|---|---|---|---|---|---|---|---|---|
| | | | *Age range* | *Subjects* | *Cross-sectional or longitudinal* | *Elicited or naturalistic language data* | *Language comprehension and/or production* | | |
| Verb agreement | BSL | Emergence and mastery | 1;10–3;0 | 1 native signing Deaf child | Longitudinal | Naturalistic (play with parents) | Production | No | Morgan et al. (2003)<br>Morgan et al. (2006) |
| Verb agreement | BSL | Emergence and mastery | 3;0–12;0 | 30 Deaf children | Cross-sectional | Elicited (picture elicitation task) | Production | No | Morgan & Woll (2003) |
| Verb agreement | DGS | Emergence and mastery | 2;2–3;4 | 2 native signing Deaf children | Longitudinal | Naturalistic (play with parents) | Production | No | Hänel (2003, 2005) |
| Verb agreement | ASL | Emergence and mastery | 1;6–3;9 | 3 native signing Deaf children | Longitudinal | Naturalistic | Production | No | Meier (1982, 1987, 2002) |
| Verb agreement | ASL | Emergence and mastery | 2–8 years | 51 Deaf children | Cross-sectional | Elicited (test) | Comprehension | No | Bellugi et al. (1995) |
| AB verb construction (incl. reference shifting and body classifiers) | BSL | Emergence and mastery | 3;0–12;0 | 30 Deaf children | Cross-sectional | Elicited (picture elicitation task & multiple-choice) | Comprehension and production | No | Morgan & Woll (2002b)<br>Morgan et al. (2002)<br>Morgan & Woll (2003) |
| Classifier constructions (whole entity classifiers, handling classifiers, SASSes) | ASL | Mastery | 4;5–9;0 | 24 Deaf native signing children | Cross-sectional | Elicited (picture elicitation task) | Production | No | Schick (1987, 1990) |

| *Linguistic Structures* | *Language* | *Emergence and / or mastery of structure* | *Methodological issues* | | | | | *Cross-linguistic comparison* | *Author(s) & year* |
|---|---|---|---|---|---|---|---|---|---|
| | | | *Age range* | *Subjects* | *Cross-sectional or longitudinal* | *Elicited or naturalistic language data* | *Language comprehension and/or production* | | |
| Classifier constructions (whole entity classifiers, hand-ling classifiers, SASSes) | Aus-lan | Mastery | 4;0–10;9 | 27 Deaf native signing children | Linked cross-sectional | 2 elicitation tasks | Production | To a certain degree with the Schick data | De Beuzeville (2004, 2006) |
| Classifier constructions (whole entity classifiers, hand-ling classifiers) | ASL, NGT | Emergence and mastery | 1;10–12;0 | 35 Deaf children | Longitudinal | Naturalistic (younger children) and elicitation | Production | No | Slobin et al. (2003) |
| Classifier constructions (movement and path, | BSL | Emergence and mastery | 1;10–3;2 and 3;0–4;11 | 1 Deaf native signing child and 18 native signing children | Longitudinal and cross-sectional | Naturalistic and elicited | Production and comprehension | Yes, e.g. Georgian | Morgan et al. (2008) |
| Spatial relations | ASL | Mastery | 4;11–9;0 (and > 12 yrs old) | 11 Deaf children | Cross-sectional | Elicited | Comprehension | Yes, with English | Martin & Sera (2006) |
| Negation | ASL | Emergence (and partial mastery) | 1;0–4;11 | 51 Deaf native signing children | Cross-sectional | Naturalistic | Production | No | Anderson & Reilly (1997) |
| Negation | ASL | Emergence (and partial mastery) | 1;6–4;11 | 16 Deaf native signing children | Longitudinal | Naturalistic | Production | No | Anderson & Reilly (1997) |

#### 2.5.1.1 Verb Agreement in DGS and Cross-Linguistic Differences

Research on verb morphology in DGS (Glück & Pfau, 1998, 1999; Happ & Vorköper, 2005) indicates that there are similar verb categories for DGS as for other sign languages (i.e., plain, agreement, and spatial verbs).

In agreement verbs, information about subject and object are indicated through the beginning and end points (loci) of a verb sign. For example, the movement from the signer to an addressee indicates 1st person subject, 2nd person object agreement. When referents are not present, loci need to be introduced by a noun or pronoun associated with that locus (Papaspyrou, von Meyenn, Matthaei, & Herrmann, 2008). Additionally, information about shape or other characteristics of an object undergoing an action (e.g., the shape of a big book handed from one person to another) can be indicated by the handshape (Glück, 2001).

In instances where agreement between subject and object is not expressed by the verb for phonetic (or pragmatic) reasons, the DGS-specific (auxiliary-like construction) of Person Agreement Marker (PAM) AUF is used (Rathmann, 2003; Rathmann & Mathur, 2002)[7]. There are different forms for this PAM in DGS, depending on whether the object is one person or a group of people (Papaspyrou et al., 2008). For example, the DGS verb HASSEN (to hate) agrees only (overtly) with the object[8]. It is produced from the signer's chin forward, using a 5-handshape, palm directed to the left side. In a sentence like *I hate you*, the verb agreement with the subject (1st person singular) is not overtly marked, but indicated through the fixed beginning point of the verb at the signer's chin and then the verb agrees with the object (2nd person singular) by the end point of the sign. But no agreement is possible with the 1st person singular as object, for example, in a sentence like *she hates me*. In this case the subject (2nd person singular) is indicated by pointing to the person, followed by the PAM from the subject to-

[7] The PAM AUF can also be used, for example, with plain verbs in DGS that are articulated on the body and thus cannot show agreement. For example, the DGS plain verb MÖGEN (to like), the PAM is the only device available to establish agreement between subject and object (Rathmann, 2003; Rathmann & Mathur, 2002).

[8] Papaspyrou et al. (2008) differentiate two types of agreement verbs for DGS: bidirectional agreement verbs, like GEBEN (to give) and FRAGEN (to ask), that *can* agree in subject and object; and mono-directional verbs that agree only in the object since they are restricted in their place of articulation. For example, the DGS sign BESPUCKEN (to spit) is produced starting close to the mouth; only the end point of the sign can vary. HASSEN also belongs in this category.

wards the signer (object) followed by the verb HASSEN. The PAM is a DGS-specific aspect that has been researched in greater depth than other aspects. PAM has also been reported in other sign languages (e.g., NGT: Bos, 1994).

Morgan et al. (2006) explored typological and modality-specific issues in the acquisition of BSL. They note that although morphological person-verb agreement uses the same principles across different sign languages, realized by movement through space, the semantic coherence between lexical items within a particular subclass of verbs that can or cannot take agreement may differ across sign languages (Morgan et al., 2006). The observation of both similarities and differences across sign languages is important (e.g., the PAM found in DGS does not occur in BSL). In addition to the DGS PAM, Morgan et al. (2006) also identify other differences between sign languages. For example, the transitive stative verb HATE can be inflected for morphological verb agreement in ASL but not in BSL, while the transitive eventive verb SUPERVISE can be inflected in BSL, but not in ASL. Cross-linguistic studies of sign languages have identified other differences in, for example, verb agreement, as in Rathmann and Mathur's (2002) comparative study of ASL, Auslan, DGS, and Japanese Sign Language.

In sum, DGS has verb agreement at its disposal, comparable to some extent to verb agreement in BSL, but different in terms of which verbs can or cannot mark agreement (Morgan et al., 2006). This is a language-specific feature that needs to be represented in an adapted DGS test. It may be expected that this category will be relatively reliable in DGS.

In the next section, studies of the acquisition of verb agreement will be presented.

#### 2.5.1.2 Acquisition of Verb Agreement

Cross-sectional and longitudinal studies using a naturalistic or experimental design indicate that the use of sign space for inflectional verb agreement morphology emerges around age 2;0 and that productive use is mastered around age 3;0 to 3;6 (Bellugi et al., 1988; Meier, 2002; Morgan et al., 2003; Morgan & Woll, 2002b) when referring to referents that are present. The ability to assign and maintain a nonpresent referent to an abstract locus starts around 4;6 to 4;11 (Bellugi et al., 1988) and its integration to verb agreement is mastered past 5 years (Bellugi et al., 1988; Morgan, 1998, 2000; Morgan et al., 2006). In those studies, the establishment and maintenance of a nonpresent referent and its integration into verb agreement is referring to narratives, which poses additional cognitive demands on the child that may influence the age at which the inflections are used (Morgan et al., 2006).

The comprehension of structures involving the assignment of nominals to abstract loci, which is important in order to successfully process and produce an agreeing verb correctly, is acquired at around 3 years for nonpresent referents (Bellugi et al., 1988). With respect to "comprehending and remembering spatial loci for referents, as well as processing verb agreement using these loci – deaf signing children evidence good comprehension of the verb agreement system by age 5" (Bellugi et al., 1988, p. 146). This relatively late comprehension may be because the methodology used acting-out, which poses additional difficulties in comprehension, since in comprehension, remembering and retrieval skills are involved.

Even when the comprehension of inflectional verb agreement is acquired by 3 years for nonpresent referents, it continues to be an important aspect of sign languages, and should thus be included in the adapted DGS test (also keeping in mind that children as young as 3 years old should be included for the standardization).

#### 2.5.1.3 Acquisition of Verb Agreement in DGS

Only one longitudinal study investigating the DGS acquisition of verb agreement is available (Hänel, 2003, 2005)[9]. Hänel analyzed the longitudinal spontaneous language production data of two Deaf children of Deaf signing parents in the age range 2;2 to 3;2 and 2;2 to 3;4, respectively, obtained in the children's home.

At the age of 2;8 and 2;4, respectively, the children used subject-verb agreement with present referents (1st person) productively. The productive use of object-verb agreement with present referents (nonfirst person) started for both children at 2;4.

The first use of pronouns referring to non-present referents (referring to subject) occurred at the age of 2;5 and 2;3, respectively, but the productive use of pronouns for nonpresent referents started at 2;8 and 2;4. This was

[9] There is another longitudinal case study by Leuninger and Happ (1997), who investigated the DGS development of a late-learning Deaf child of hearing parents who was first educated orally before having access to DGS through weekly home instruction in DGS from 3;7. Only development between 3;7 and 4;11 was observed. The findings for this child with late first exposure are not included in this review.

followed by the first use of subject-verb agreement form with nonpresent referents at the ages of 2;8 and 2;3/2;4, respectively. The first production of pronouns for nonpresent referents (referring to object) started at 2;8 and 2;3, respectively, but productive use did not occur before 2;8. The first object-verb agreement forms with nonpresent referents appeared at 2;8 and 2;3, respectively. Hänel (2005) explained the relative earlier productive use of verb agreement for nonpresent referents in DGS as compared to ASL studies as the result of methodological differences. The DGS data were drawn from spontaneous language production, whereas the ASL data were mostly drawn from elicited picture description tasks[10]. Hänel (2005) concluded that the results of her study "when compared with analyses of ASL children, [...] clearly coincide in one important point, namely that the DGS children as well as the ASL children acquire the underlying agreement mechanism during the same age span" (p. 229).

#### 2.5.1.4 Complex AB Verb Constructions

Complex AB verb constructions constitute a subcategory of normal agreement verbs in BSL and ASL (e.g., Morgan & Woll, 2003; van Hoek, O'Grady, & Bellugi, 1987). Complex AB verb constructions "are depictions of actions performed by an individual on a specific body part of another individual" (Morgan & Woll, 2002b, p. 266), such as in sentences like *the boy taps the girl on the shoulder* or *the girl combs the boy's hair*. The semantic information is realized using the manual and non-manual channel. Complex AB verb constructions involve an event with two participants. The signer is required to locate two referents in sign space through indexing. The main verb is inflected from two shifting perspectives: the first perspective shift indicates agent and action (e.g., *boy taps* or *girl combs*); the second perspec-

[10] The previously reviewed ASL and BSL acquisition studies on verb agreement (e.g., Bellugi et al., 1988; Morgan, 2000; Morgan et al., 2006) indicate a later acquisition of verb agreement referring to nonpresent referents (past 5 years old) than the DGS study by Hänel (2003, 2005). A plausible explanation – additionally to the methodological issue already mentioned by Hänel (2005) – is that in the ASL and BSL studies on verb agreement the criteria on the mastery of verb agreement with non-present referents is defined as the establishment and maintenance of abstract loci for pronouns, which is associated with longer stretches of discourse than compared to single sentences (e.g., narrative production of a child). This poses additional cognitive demands to the child and may influence the age at when the child really uses verb agreement with non-present referents, thus providing a possible explanation for the difference in the developmental timetable.

tive shift indicates the action, the experiencer, and the body part affected (e.g., *taps girls shoulder* or *combs boy's hair*). Another option is that the signer does not locate the two referents by indexing, but uses the whole entity classifier for *person* to indicate the two referents, which are located on the right and left side of the sign space (Morgan & Woll, 2002b). The non-manual marker of perspective shift is produced twice, as compared to normal verb agreement pattern. "The three arguments encoded by the AB verb cannot be mapped through a single verb, and in the AB construction, the extra argument of the affected body part is mapped onto the B-part of the verb" (Morgan & Woll, 2002b, p. 267). AB verbs encode three arguments: the agent, the patient and the affected body part (Morgan et al., 2002).

> An important part of the structure is the non-manual marker of shifted reference.…The non-manual markers are produced simultaneously with the onset of the perspective shift prior the verb. (Morgan et al., 2002, p. 663)

AB verb constructions require an exchange of referent location in sign space. They also require the use of body classifiers, such as the affected body part involved (e.g., the shoulder in *the boy taps on the girl's shoulder*). This definition of the term *complex AB verb constructions* will be used for working purposes within the present study.

Research on DGS that specifically addresses complex AB verb constructions could not be found. In theory, this construction should also be available to DGS, since the different components of manual and non-manual channels are available (as described in the next section on the acquisition of AB verb constructions). It has been argued above that inflectional verb morphology (agreement verbs) is also available in DGS, as is role shift (Papaspyrou et al., 2008). It can be assumed that AB verb constructions exist in DGS, but their exact linguistic status (i.e., whether they constitute a discrete subcategory of verbs) is not known.

#### 2.5.1.5 Acquisition of Complex AB Verb Constructions

The basis for acquiring complex AB verb constructions is the prior mastery of inflectional morphology in sign space (i.e., the use of verb agreement), which is mostly acquired by age 3;0 to 3;6 for present referents (Morgan et al., 2006). An AB verb construction poses additional complexity in the need for simultaneous organization of manual and non-manual channels. The full acquisition of non-manual features often follows the acquisition of manual features in other areas of sign languages acquisition (e.g., for negation: Anderson & Reilly, 1997). Most importantly children need to learn how to map the conceptual knowledge of the agent, patient, and affected body part onto the two parts of the verb (Morgan & Woll, 2002b).

First, the acquisition of the structures of complex AB verb constructions in comprehension and production will be discussed, followed by the acquisition of the semantics of AB verb constructions. In a cross-sectional study using an experimental design, the earliest age at which comprehension of complex AB verbs constructions has been reported in BSL is 3;2 in a study analyzing two BSL sentences using AB verb constructions (Morgan et al., 2002; Morgan & Woll, 2002b). Children in this study aged 3;2 to 12;0 (*N* = 30) were grouped into three age bands. Between 3;2–5;11 children show a gradual emergence of partial comprehension (40% of children) of AB verb constructions, with increasing comprehension in children ages 6;0–8;11 (60% for the first target sentence, 80% for the second sentence), with nearly complete mastery in children aged 9;0–12;0 (90%) (Morgan et al., 2002; Morgan & Woll, 2002b).

Production of these sentences followed a different time course. Sentences requiring the production of single agreement verbs such as *the mother gives the child a book* were correctly produced by 100% of children aged 9–12 years, but the complex AB verb construction was correctly produced by only 70% of these children, indicating the higher morphological complexity of AB verb constructions. The results of the study on complex AB verb construction also showed a clear developmental trend from no knowledge in the production of AB verbs constructions in the children of the youngest group (3;2–5;11), to 40% correct in the middle age group (6;0–8;11), to 70% in the oldest group (9;0–12;0), which was still below adult performance.

As for the acquisition of the semantics in the complex AB verb constructions, occasional examples were found of the A-part only for the earliest stage of development of AB verb constructions (around age 3;6). Some children "produce a serial ordering of the thematic relations" (Morgan & Woll, 2002b, p. 273), which may reduce complexity for the child. This is followed by a strong tendency to use the B-part (patient receiving action) only, with partial or full production of just the B-part in 90% of the youngest children, 40% of the middle group, and only 10% of the oldest children (Morgan &

Woll, 2003), which indicates a tendency to systematically omit an argument (i.e., the agent). The child attempts to map the three thematic roles of the event (agent, patient, and affected body part) onto a single transitive verb frame rather than distributing the event across the complete AB verb construction. The tendency to use just the B-part suggests an (over)generalization of the normal inflectional verb agreement pattern. Later in development, the A and B parts are used together (40% in children 6;0–8;11 and 70% in children 9;0–12;0), but children still struggle with non-manual features (Morgan et al., 2002; Morgan & Woll, 2002b), and in some children (from 6;0 onward) the A-part is initially expressed with a lexical marker which is gradually replaced by the adult-like form of the AB verb constructions from 9;0 onwards (Morgan et al., 2002). The acquisition of the required non-manual morphology in AB verb constructions is related to the development of the productive use of the referential shift, which starts at the sentential level around age 5 (Morgan, 2002; van Hoek et al., 1987).

AB verb constructions show a clear age-related pattern in both comprehension and production. For the purpose of this test, the comprehension of AB verb constructions in BSL is useful, as it shows a clear developmental trend, starting around the age of 3;2 and completed by the age 9–12 years. This is an important issue for the adaptation of the DGS test.

The acquisition studies of AB verb constructions in ASL and BSL are the only ones that address this issue, and realization is similar in the two languages.

In the next section, a review of classifier constructions in DGS and cross-linguistic differences, followed by the acquisition of classifier constructions will be presented.

#### 2.5.1.6 Classifier Constructions in DGS and Cross-Linguistic Differences

*Spatial verbs with whole entity classifiers* in DGS show similar formational patterns to those described for BSL and other sign languages. Spatial verbs can be further divided into two (sub)-categories: (1) locative verbs (e.g., *a book is on the table*); and (2) verbs of motion (e.g., *a car drives up the hill*) (e.g., Glück, 2001; Glück & Pfau, 1997a, 1997b, 1998; Happ & Vorköper, 2005; Perniss, 2007). In sentences with spatial verbs, the lexical noun precedes the verb with a whole entity classifier (e.g., CAR CL-DRIVE-UP-HILL).

This review of the literature on spatial verb morphology in DGS suggests that it has similar linguistic features at its disposal as does BSL. However, different sign languages may use different handshapes to classify whole entities. For example, in ASL an Upright-3 handshape represents vehicles, whereas DGS and BSL use a Flat-B handshape to represent vehicles (Boyes Braem, 1995; König, Konrad, & Langer, in press; Sutton-Spence & Woll, 1999). In relation to other classes of objects, the choice of handshape is less restricted, depending on the shape and the size of the object to be represented (König et al., in press).

*Size and shape specifiers*: Like BSL, DGS has size and shape specifiers (SASS) that describe the size, shape, pattern, and dimensions of a noun/referent (Glück, 2005; Happ, 2005). SASSes in DGS can function in a comparable way to adjectives in spoken languages (e.g., Glück, 2001; Glück & Pfau 1997a, 1997b). The movement of a SASS does not refer to the movement of a referent in space; instead, the movement is used to draw a referent's size, shape, and dimensionality.

The review of the DGS literature suggests that it has signs that function like SASSes in BSL and other sign languages. Since SASSes describe size and shape, cross-linguistic differences are likely to be small. As they refer to the salient features of a referent (i.e., to represent stripes on a shirt), it is likely that similar handshapes will be used in different sign languages since the choice of handshape is motivated by the referent's size and shape (König et al., in press).

*Handling classifiers*: It has been reported that like BSL, DGS also has handling classifiers (Glück, 2001; Glück & Pfau 1997a, 1997b, 1998; Happ & Vorköper, 2005). Handling classifiers convey information about how an object can be manipulated by an animate being. The choice of handshape reflects properties of the object that is manipulated. For example, in the sentence *the man gives the woman flowers*, the handshape provides information about the shape of the part of the flowers that is held (the stems as a thin object) and the movement represents the act of giving. In the sentence *he drinks glass of water*, the handshape represents holding a cylinder with a small diameter.

A review of the literature suggests that DGS also makes use of handling classifiers. These are likely to be similar across sign languages, since the representation of the manipulation of an object is language independent[11]

[11] Although the manipulation of a certain object might be language independent, it might not be culture-independent. For example, the use of a spoon will vary depending on how people in a culture actually hold and use spoons.

and the choice of handshape will depend on the part of the object being manipulated.

In sum, the literature on DGS indicates that whole entity classifiers, SASSes, and handling classifiers also are found in the target language of the adapted test, which is an important prerequisite for the adaptation, and thus it may be expected that this category will be relatively reliable in DGS.

Certain handshapes used for whole entity classifiers (e.g., to represent vehicles) are conventionalized for a certain language and so will differ across sign languages. However, there are strong cross-linguistic similarities for SASSes and handling classifiers (Schembri, 2003). These cross-linguistic or "universal" structures in sign languages result in a "relative transparency in the way that some types of objects or events in the world are linguistically encoded" (Sandler & Lillo-Martin, 2006, p. 509). This is an important issue for the adaptation of sign language tests. If certain structures are likely to be similar across sign languages, then tests involving these items are more suitable for adaptation into other languages, even when the target language is not fully documented.

#### 2.5.1.7 Acquisition of Classifier Constructions

In this section, first an overview on the development of spatial concepts in the cognitive domain (Piaget & Inhelder, 1956) and its link to (spoken) language acquisition will be presented, followed by a review of the acquisition of classifier constructions[12] across sign languages.

*The development of spatial concepts and spatial language*: Children pass through different stages in the development of spatial concepts. Once representational thought has been developed, children start to develop firstly topological spatial concepts followed by projective and Euclidean spatial concepts (Piaget & Inhelder, 1956). These spatial concepts express the relationships of two objects to each other.

Topological spatial concepts include concepts like *next to*, *on*, *in* or *between*. Projective and Euclidean spatial concepts are acquired in tandem including concepts like *in front*, *back*, *right*, and *left* (Liben, 2006). Projective spatial concepts refer to the child's point of view, that is, the child's under-

[12] Most acquisition studies group spatial verbs with whole entity classifiers, handling classifiers, and SASSes together. Therefore, for working purposes, the term *classifier constructions* will be used from here on to refer to whole entity classifiers, handling classifiers, and SASSes.

standing the effect of viewing position on the appearance of an object or group of objects from a different perspective (e.g., Liben, 2006). Euclidean concepts can be best called "abstract spatial system" concepts because they provide the structure by which locations and objects are represented in an abstract, stable, and general three-dimensional system (Liben, 2006).

The best-known experimental task regarding the point of view of children is the *Three Mountain Task* by Piaget (Piaget & Inhelder, 1956). Children are shown a three-dimensional model of three mountains with three different peak heights. Children are asked to indicate (by selecting one of several pictures) how the mountain would look when someone would sit, for example, across the table from the child. This task can determine if the child can appreciate that another person's perspective is different from his/her own perspective and thus overcome egocentrism. Children in the preoperational stage (up 6;6 years old) use their own perspective in projective relations. With the appearance of the operational stage (7 to 9 years old), a more progressive differentiation between points of view and certain projective relations are formed, with the concepts of *before* and *in front* differentiated before *right* and *left* (Hart & Moore, 2005). With the final equilibrium of concrete operations (around 9 to 10 years) "the two schemas are inter-coordinated and the child masters a comprehensive coordination of viewpoints completely independent of his own view" (Hart & Moore, 2005, p. 267). Euclidean relations are developed in the stage of formal operations (about 11 or 12 years old) and onward (Hart & Moore, 2005).

Cross-linguistic studies on the acquisition of spatial concepts in children suggest that their first spatial words are applied to the same kinds of (spatial-related) events; for example, putting things *into* containers and taking them *out*, piling things *up* and knocking them *down* (Bowerman & Choi, 2003). In English, children first acquire spatial terms relating to containment (*in*, *out*), accessibility (*open*, *close*, *under*), contiguity and support (*on*, *off*), verticality (*up*, *down*), and posture (*sit*, *stand*), followed by words of proximity (*next to*, *between*, *beside*), and later words for projective and Euclidean relationships (*in front*, *behind*) (Bowerman & Choi, 2003). Bowerman and Choi (2003) state that "this sequence of development is consistent with the order of emergence of spatial concepts established through nonlinguistic testing by Piaget and Inhelder (1956)" (p. 391).

However, languages differ as to how they encode spatial concepts in terms of domain (for example, the domain of *ON*) and segment these spatial concepts into different categories (Genter & Bowerman, 2008). For example, Bowerman and Choi (2003) discuss differences of spatial categories in English and Korean. Whereas English uses the concept of *put in* as a unified category of containment events, Korean speakers subdivide: "tight-fit containment events like putting a book into an exactly matching box-cover, described with *kkita*, are treated as a different class of actions from loose-fit containment events like putting an apple into a bowl or a book into a bag, described with *nehta*" (Bowerman & Choi, 2003, p. 392).

Genter and Bowerman (2008) show that in some languages, like English, *on* encodes several concepts compared to Dutch which uses at least three different terms: *op*, *om*, and *aan*[13]. Spatial concepts are encoded differently across languages, having different and also overlapping categories.

The different encoding of spatial concepts in different languages affects language acquisition (e.g., Choi & Bowerman, 1991). Additionally, the encoding in one language may be realized with a preposition and in another language by means of case endings, or verb form (Johnston & Slobin, 1979; Slobin, 1973). Choi and Bowerman (1991), concluding their study on the different developmental timetable of Korean and English, state that

> the meanings of children's early spatial words are language specific. This means that language learners do not map spatial words directly onto non-linguistic spatial concepts, as has often been proposed, but instead are sensitive to the semantic structure of the input language virtually from the beginning. (pp. 117–118)

Briefly, the cognitive or non-linguistic development of spatial concepts proceeds through different stages as proposed by Piaget and Inhelder (1956). The *order* of development of spatial semantics parallels the development of cognitive development (e.g., Bowerman, 1996) although languages differ in how they segment and encode spatial concepts, and this has an impact on their development. In the case of the present study, it is of interest to see how Deaf children learn to comprehend and produce topological, projec-

[13] Whereas English uses the word *on* to refer to topological spatial relations of *cup on table*, *band-aid on leg*, *apple on tree*, and *ribbon on candle*, Dutch encodes these spatial relations using three different words (Genter & Bowerman, 2008).

tive, and Euclidean spatial relations in a sign language that is encoded in classifier constructions.

*Classifier constructions*: Within spatial verbs, whole entity classifiers are verb or predicate constructions in which the handshape substitutes for the overall shape of the class of nouns to which the action refers. For example, the Flat-B handshape might be used to refer to vehicles or the Upright-1 handshape to refer to a person. The whole entity classifier can move around in space or can be located somewhere, mirroring movements and locations of the real world. Handling classifiers represent how a hand holds or handles a class of referents, and the action of the signer's hand indicates the manipulation of the referent, such as *take a book from the shelf*. In size and shape specifiers (SASS), movement and handshape are used to outline the size or shape of a referent or features of that referent. The movement does not represent movement of the referent in space, but rather describes its features and their dimensions, such as the stripes on a shirt[14].

The emergence of classifier constructions for the description of motion and location events were reported to appear early in a Deaf native signing child acquiring BSL (from 2;0 onwards) in a longitudinal study collecting spontaneous language data in the child's home (Morgan, Herman, Barrière, & Woll, 2008)[15]. Before 2;0 whole body depictions were used to describe, for example, movements, "such as lifting the arms for 'jumping' and moving the hands forward to describe 'falling'" (Morgan et al., 2008, p. 8). But only figure and path components were encoded. Between 2;0 and 2;6 these gestural forms disappeared and were replaced by classifier constructions. The child also provided more

> event information about ground, path or manner, using either finger tracing, real-world objects or the physical ground itself. In several utterances, quite elaborate manners of movement and paths were expressed through tracing of an index finger, e.g., POURING, ZIGZAGGING, PIROUETTING, OVERTAKING and CROSSING-OVER. Each of these motion and location

[14] For a detailed discussion of the different models and terminologies that are used when referring to classifier constructions, see Schembri (2003).

[15] Studies have been undertaken on the acquisition of classifier constructions for nearly 30 years (e.g., Kantor, 1980; Supalla, 1982), but due to changes in theoretical approaches and categorization of these structures, the review will be based on more recent studies.

> descriptions was preceded by a sign for the nominal CAR, PLANE, MAN, etc. but the child did not combine a handshape classifier for figure with the movement of the hand. (Morgan et al., 2008, p. 8)

In this period, most errors arose from the selection of a wrong handshape in verbs of motion and location, for example, the use of the extended index finger handshape to refer to a plane and helicopter (Morgan et al., 2008), indicating a different developmental timetable in classifier constructions of handshapes on the one hand and motions and locations on the other. Even with an increasing use of handshapes and combinations of movement and location between 2;6 and 3;0, the selection of incorrect handshapes continued past 3;0. This means correct productive use lagged behind comprehension. Between 2;0 and 3;0 the productive use of movement started to emerge (Morgan et al., 2008), but there was no clear indication of productive use of location before 2;6.

Examples of the gradual emergence of classifier constructions have also been reported for both Deaf children of Deaf parents and Deaf children of hearing parents in the US and the Netherlands, in both longitudinal studies collecting data in naturalistic settings with younger children, and in structured settings retelling a story with older children around age 12 years old (Slobin et al., 2003). Handling classifiers[16] are acquired relatively early, having been reported in a Deaf child of Deaf parents aged 1;10 acquiring ASL. The child produced an utterance describing the placement of a thin, flat object onto another object, although without referring to the depth or thickness of the object with the meaning of *putting a book onto another book* (Slobin et al., 2003). However, emergence of such constructions in other children has been reported to occur at a later age. For example, *pushing a buggy* was observed in a Deaf child of hearing parents at age 2;6 acquiring

[16] Slobin et al. (2003) define two types of handling classifiers or "handle handshape units": *manipulative handle handshapes* and *depictive handle handshapes*. In manipulative handle handshapes, the movements represent the movement of the hands manipulating an object that is referred to. Depictive handle handshapes are more demanding "than manipulative handles, in that they require the learner to choose an appropriate handshape for representing the salient dimensions of the object to be handled, rather than representing the manipulating hand itself" (Slobin et al., 2003, p. 280). For the purpose of this study, the acquisition of these two types of handles will not be treated separately (depictive handles are not represented in the original BSL and the adapted DGS test).

NGT (Sign Language of the Netherlands); *unzipping a suitcase* was observed in a Deaf child of Deaf parents at age 2;5 acquiring ASL. In these two examples the children omitted the ground handshape.

The emergence of handshapes in spatial verbs referring to vehicles (as whole entities) is acquired later (Slobin et al., 2003): one Deaf child of Deaf parents acquiring NGT at age 2;9 signs CAR as a lexical sign first, followed by a modulating of the "2-handed CAR into a 1-handed VEHICLE entity that serves as ground for the LIGHT [of an ambulance] indicated by the other hand" (Slobin et al., 2003, p. 282). They note that two-year-olds have problems coordinating two handshapes "when one represents a type of figure and the other a type of ground, or two figures in relation to one another" (Slobin et al., 2003, pp. 283–284).

No data were found on the emergence of SASSes in Deaf children from spontaneous data. Slobin et al. (2003) suggest that a possible explanation might be that elicited tasks provide a better context in which to use SASSes.

Morgan et al. (2008) also investigated the comprehension of movement and path descriptions and relative locations in a BSL sentence comprehension task in 18 native signing children age 3;0 to 4;11 (cross-sectional). The comprehension of classifier constructions representing movement and path descriptions and relative locations increased from 3;0, but the slow development of comprehension of projective and Euclidean spatial relations (e.g., *in front*, *behind*, *right*, *left*) suggests that comprehension of BSL motion and location sentences was far from complete at 5;0 (Morgan et al., 2008). Comprehension of movement and path descriptions in BSL classifier constructions appears to be easier than comprehension of (relative) locations between 3;0 and 4;11. Morgan et al. (2008) conclude that items requiring reversal in perspective were the most difficult, especially *right-left* relations: 3-year-olds scored below chance (25%) on *behind*, *under*, *in front*, *bottom-left*, *inside-right*, and *top-left* (Morgan et al., 2008).

The acquisition of spatial language is difficult, and it is acquired relatively late. A picture-matching comprehension task of spatial relations in ASL with three groups (cross-sectional) revealed that in children 4–9 years, children > 12 years, and adults, the comprehension of spatial constructions such as *above* and *below* (which do not require mental rotation) pose fewer problems than the comprehension of spatial relations requiring mental rotation, such as (1) *in front*, (2) *behind*, (3) *right*, (4) *left*, (5) *towards*, and (6) *away* (Martin & Sera, 2006). All these constructions involve figure and ground; the spatial relations of (1) through (4) are static, whereas (5) and (6) also involve a path movement. Adults performed better on these spatial relations than the younger group of children. All spatial relations involving mental rotation show an age effect, with all children worse than adults. Rotated constructions in ASL posed greater difficulty than the same spatial constructions in English, especially among children. The scores of children aged 4–9 were best with *above,* with a descending order of correct responses for *below*, *in front*, *behind*, *towards*, *right*, and with *left* and *away* scoring lowest, suggesting that they are the most difficult to acquire. Comprehension of these concepts is acquired relatively late, as also found by Morgan et al. (2008). This is also in agreement with development of spatial concepts in the cognitive domain (Piaget & Inhelder, 1956).

Studies of the mastery of classifier constructions in Auslan (de Beuzeville, 2006) and ASL (Schick, 1987, 1990) indicate a hierarchy in the development of these constructions in native signing Deaf children aged 4;0 to 10;9 years (de Beuzeville: 4;0–10;9; Schick: 4;5–9:0). The studies produce slightly different results, but the main conclusions are the same: handling classifiers are acquired first, followed by SASSes, and then whole entity classifiers[17] (de Beuzeville, 2006; Schick, 1987). These studies did not focus on younger children (< 4 years old), where these structures emerge (as does Slobin et al., 2003), but rather focused on the developmental path from > 4 years onward to its mastery.

In the cross-sectional study of ASL (Schick, 1987) using one elicitation task, there was a clear developmental progression in the acquisition of the use of space, but no clear age-related differences were found for handshape and movement. There was a trend for handling classifier and SASS handshape production to improve with age, but this was not apparent for whole entity handshapes (Schick, 1987)[18]. Children had the greatest difficulty in

[17] Schick (1987) reports the sequence of acquisition of adult-like production of the different classifier constructions as handling classifiers and SASSes acquired at the same age (across the three age groups, more handling classifiers were produced than SASSes, but the difference was not statistically significant), followed by spatial verbs with whole entity classifiers.

[18] Fish, Morén, Hoffmeister, and Schick (2003) conducted a cross-sectional study on the acquisition of handshapes in classifier constructions in ASL using the Real Object production task of the ASL Assessment Instrument (Hoffmeister, 1999) with 144 Deaf children of Deaf and hearing parents aged 3-12 years old. The authors conclude that "handshape complexity *alone* seems to make the wrong predictions for VOL [verbs of location] production, and the

using space to locate objects relative to other objects in complex arrangements; for example, where the child had to "construct a Euclidean reference system that is independent of perspectives. These concepts are acquired relatively late" (Schick, 1987, p. 93). The use of space was strongly associated with developmental progression. Additionally, there were significant differences in complexity of classifier constructions. Across handling classifiers, SASSes, and whole entity classifiers, simple constructions were produced more correctly than complex constructions. Schick also found that these three classifier constructions were not fully mastered even by 9;0 (Schick, 1987).

In her cross-sectional experimental study on the acquisition of Auslan, de Beuzeville (2004, 2006) used two elicitation tasks that had been used in two previously conducted ASL acquisition studies. One task was based on a study by Supalla (1982) and the other elicitation task was based on the reviewed study by Schick (1987). De Beuzeville's study focused on the acquisition of whole entity classifiers, handling classifiers, and SASSes in native signing Deaf children ages 4;0 to 10;9. The results indicated the following developmental order: (1) handling classifiers; (2) SASSes; followed by (3) whole entity classifiers. The following lists the order of mastery of the three parameters in de Beuzeville (2006): for handshapes, handling classifiers > whole entity classifiers > SASSes; for movement across the three classifier constructions (only 4-year-olds differed by movement), handling > whole entity > SASS; for children older than 4 years there was no statistically significant difference between movement types; the order of acquisition for location was handling > SASS > whole entity.

The order of acquisition of handshape, location, and movement across the three classifier constructions differed in the Schick study (1987, 1990): for handshape, the order of acquisition was whole entity classifiers > SASSes > handling classifiers; for movement, the order was whole entity classifiers > SASSes = handling classifiers (with no statistically significant difference between SASSes and handling classifier movements); for location, the order was handling classifiers > SASSes > whole entity classifiers.

The results in more detail of de Beuzeville's study (2006) are:

---

phenomenon of classifier acquisition in ASL appear to be more complex and morphologically driven than thought" (Fish et al., 2003, p. 262).

> The children in the study used adult-like handshape, movement and location forms for handling DVs [handling classifiers] 90% of the time at age 4. Tracing DVs [SASS] took longer to master, reaching 90% at age 6, before dipping at 7 [across handshape, movement, and location] and reaching 86% again at age 8. Modelling verbs [whole entity classifiers] were mastered last. Indeed, in the data for this project, even the 10-year-olds did not yet have adult mastery of these signs, scoring accurately only 75% of the time. (p. 192)

While the youngest children (4–5 years) were able to use handling classifiers, SASSes, and whole entity classifiers, they sometimes used a whole body enactment to animate a whole entity classifier or handling classifier. Sometimes they also just used a lexical sign or did not respond[19]. Four- and five-year-olds used lexical signs about twice as frequently as older children. Frequently, the integration of figure and ground handshapes in the classifier constructions was not observed: 4–5 year-olds omitted the ground handshape in 40% of all complex spatial verbs. However, de Beuzeville notes that 4-year-olds produced many items that included a ground handshape, so the feature may be in the course of emergence at that age (de Beuzeville, 2006). Children in this age group also make handshape errors and have the least control over the use of sign space.

Children aged 6–8 are in the middle of the developmental course. In this age group, children responded in

> similar ways as younger participants, but the frequency of certain phenomena was different. That is, at these ages the children were signing fewer whole body enactments and frozen signs and responding more often overall. They more often used ground handshapes in complex modeling verbs [whole entity classifiers] than the younger children, although still exhibited considerable troubles doing so. (De Beuzeville, 2006, p. 200)

In the oldest group (9–10 years), there is a sharp drop in the use of whole body enactment and their production resembled that of adults in all para-

[19] Possible explanations are (1) inattention of the children, (2) lack of understanding of the task, or (3) avoiding the task because it was too difficult (de Beuzeville, 2006).

meters. Children did not produce all handshapes with adult-like accuracy, but there was a clear increase in the use of the target handshape in this group. Among 10-year-olds, the ground handshape was rarely omitted (by only 10%).

Both studies (de Beuzeville, 2006; Schick, 1987) suggest a developmental hierarchy for handling classifiers, SASSes, and whole entity classifiers in Deaf native signing children aged 4;0 to 10;9 across sign languages. Even the oldest children in both studies (> 9 years) had not yet acquired adult-like production of whole entity classifiers. A further study, on the acquisition of classifier constructions in Brazilian Sign Language (LSB; Bernardino, 2005) also found that a Deaf child with Deaf parents had not yet completely mastered the classifier system by age 9.

Morgan et al. (2008) claim that children learn the conventions for specifying the locations of one object with respect to another using prospective and Euclidean principles at around 11 or 12 years. Slobin et al. (2003) also note that Deaf children in late preschool and early school age have difficulty incorporating the ground classifier in classifier constructions, with older school-age children having a more sophisticated understanding of the different functions of ground classifiers. Ground can also be considered in terms of scale when a figure moves or is located relative to a ground (i.e., figure and ground need to correspond to one another) (Slobin et al., 2003, p. 290 for more details). A 5-year-old child they studied had not mastered this construction, using the wrong handshape to indicate the ground and not providing an appropriate scale. A 12-year-old child, who was a non-native but skilled signer, could successfully choose figure and ground classifiers corresponding with one another appropriately in terms of scale. Figure is generally acquired before ground (Tang, Sze, & Lam, 2007). The use of both figure and ground in complex constructions may emerge around age 4–5 years (de Beuzeville, 2006), but mastery is achieved much later, at around 11–12 years (Morgan et al., 2008; Slobin et al., 2003).

The developmental path for classifier constructions can be charted from the emergence of the earliest incomplete handling classifiers in native signing children around 2 years of age, the emergence of the first whole entity classifiers around 2;6 (no data is provided for the emergence of SASS constructions), and the near complete mastery of handling classifiers at 4–5 years[20], SASSes at age 6–8, and whole entity classifiers from 10–11 years onward across different sign languages (LSB: Bernardino, 2005; Auslan: de Beuzeville, 2004, 2006; BSL: Morgan et al., 2008; ASL: Schick, 1987, 1990; ASL and NGT: Slobin et al., 2003). Most studies focus on the acquisition of production of these constructions, not comprehension. The few studies that are concerned with the development of comprehension of these constructions (Martin & Sera, 2006; Morgan et al., 2008) indicate that native signing children start to comprehend movement and path descriptions and relative locations with figure and ground at around age 3;0, (Morgan et al., 2008), but still struggle with full comprehension of more complex static spatial relations such as *behind*, *in front*, and *right-left distinction* at around 9 years of age (Martin & Sera, 2006), with complete comprehension not mastered until after 9.

These studies provide evidence from different sign languages relating to the emergence and mastery of aspects of classifier constructions. Although a complete picture is lacking, some pieces of the puzzle are available. The timetable and developmental hierarchy is similar across sign languages and provides a basis for the adaptation and development of sign language tests.

In the following section, number and distribution will be discussed.

#### 2.5.1.8 Number and Distribution in DGS and Cross-Linguistic Differences

Number and distribution are ways in which sign languages express plurals, with a similar meaning to plurals in spoken languages, but the morphological forms in sign languages can be more complex (Sutton-Spence & Woll, 1999).

---

[20] The results of Schick (1987) differ from those of de Beuzeville (2006). For example, Schick found that the oldest children (7;5-9;0) scored only 58% correct on handling classifiers, indicating that the acquisition of classifier constructions was not yet completed, while de Beuzeville (2006) reported mastery of handling classifiers between the age of 4-5 years. These differences also relate to SASSes. The differences in the acquisition timetable can stem from different methodological criteria that were defined for "correct" adult-like production in both studies. In the de Beuzeville (2006) study, the Schick task was also filmed with five native signing Deaf adults as controls. The adult responses were accepted as the targets and used as the means of comparison as "correct" adult-like production. De Beuzeville argues that she uses less strict criteria for "correct" adult-like production and thus accepted a wider range of options as correct forms than Schick (1987) did (L. de Beuzeville, personal communication, April 23, 2009).

Perniss (2001) describes the various ways in which DGS plurals can be realized, depending on the context. In order to express the exact quantity of a nominal sign, a numeral sign ZWEI (two) can be produced before the nominal sign. Quantifiers such as VIELE (many) can be produced before or after the noun, although location after the noun indicates special emphasis (Perniss, 2001). A simple unspecific plural of a nominal sign is realized by a repetition of the sign, normally between two to four times (Perniss, 2001) (e.g., the singular form is HAUS (house) and the plural form is HAUS++). The repetition of the sign occurs at the same location in sign space (Perniss, 2001). In isolation, the sign is only repeated twice, but in context, the sign can be repeated three or even four times. Perniss analyzes the repetition of a nominal sign, which includes a sideways movement, as dual rather than an unspecific plural (Perniss, 2001). Another option to express plurality (specific or unspecific) is to sign a nominal (e.g., HAUS), followed by a numeral sign or a quantifier, and then to index the nominal in sign space (in this case the indices of the nominal do not refer to the specified location of the referent) (Perniss, 2001).

Pfau and Steinbach (2005, 2006) state that there exist three types of noun pluralization in DGS. These three types of noun pluralization are (1) sideward reduplication, (2) simple reduplication, and (3) zero marking. They involve three types of place of articulation (lateral and midsagittal plane, body-anchored) and two types of movement (simple and complex) (Pfau & Steinbach, 2006). Based on these five features (place of articulation and type of movement), four different kinds of nouns can be distinguished in DGS: (1) body-anchored nouns (B-nouns) such as FRAU (woman); (2) non-body-anchored nouns with complex movements (C-nouns) such as AUTO (car); (3) non-body-anchored nouns with simple movement in midsagittal space (M-nouns) such as HAUS; and (4) non-body-anchored nouns with simple movement in lateral space (L-nouns) such as KIND (child) or PERSON (person).

B-nouns are body-anchored nouns where reduplication is not possible. C-nouns are two-handed signs produced in the midsagittal plane with a complex movement, such as AUTO or FAHRRAD (bike) (Pfau & Steinbach, 2006). The only possible form of plural marking for such signs is zero marking (i.e., no reduplication of the noun sign). M-nouns are signed in the midsagittal plane, and are mostly two-handed signs (Pfau & Steinbach, 2006) produced with symmetrical movement. Pluralization is realized by reduplication, but sideways reduplication is not possible, and only a simple reduplication is used (e.g., the nouns are repeated twice at the same place of articulation, such as HAUS++). L-nouns are one-handed signs that are produced – depending on the handedness of the signer – on the right or left side. The pluralization is realized by a sideways reduplication (e.g., the base form is repeated twice, such as PERSON++).

B-nouns do not occur in the item pool of the BSL test and are therefore not in the adapted DGS test. One M-noun (BALL, Item 47) is in the DGS item pool, representing a singular form; thus, this form of pluralization in DGS cannot be applied to that item. The plural of M-nouns is slightly different in BSL and DGS. Whereas in DGS only a simple reduplication is possible (Pfau & Steinbach, 2005, 2006), BSL allows a sideways reduplication by adding a bound plural morpheme (Sutton-Spence & Woll, 1999). In DGS, sideways reduplication is used to indicate dual, not an unspecific plural (Perniss, 2001). No M-nouns in plural form are represented in the test. L-nouns are not represented in the original BSL item pool.

C-nouns like APFEL (apple) or AUTO are part of the item pool, but only in combination with the quantifier VIELE (Items 1 and 41). Noun + quantifier constructions are similar in BSL (Sutton-Spence & Woll, 1999) and DGS (Perniss, 2001; Pfau & Steinbach, 2005, 2006) (i.e., a quantifier followed by a noun). But following Perniss (2001), quantifiers can also be produced after the noun for special emphasis. Heßmann's corpus (2001b) indicates that the quantifier is also produced more often in the pre-nominal position. There are two items in the original BSL item pool with a quantifier which will be adapted to DGS.

Happ and Vorköper (2005), Perniss (2001), and Pfau and Steinbach (2005) report on other ways of expressing the plural. For example, by producing the lexical noun first, followed by a reduplication of the classifier handshape, similar to the description of spatial verbs with whole entity classifiers in DGS. The plural can be produced in two different ways: by reduplication of CL++ indicating that cars are parked side by side, or with a tracking movement indicating a row of cars (Perniss, 2001). These signs do not only provide number information, but also information about orientation and location in space (i.e., there is an overlap with the category of spatial verb morphology where two items belong to two categories: spatial verb morphology, and number/distribution). One exception is the classifier Upright-4 handshape representing a queue of people. This is intrinsically plural in meaning and thus does not need to be reduplicated in order to express plural. Summarizing these findings, it suggests that similar structures are used in British Sign Language and DGS to express number and distributive aspect, and thus it may be expected that this category will be relatively reliable in DGS.

DGS and BSL share features for expressing number/plural in nominal signs (e.g., number sign preceding the nominal sign, use of classifier constructions in spatial arrangements), but they differ in other features, such as the reduplication at a single location of nouns like HAUS in DGS and sideways reduplication of similar signs in BSL. The status of the position of the quantifier VIELE in DGS is unclear.

#### 2.5.1.9 Acquisition of Number and Distribution

Few studies explicitly address the acquisition of number and distribution in Deaf children. This tends to be done as part of research on the acquisition of classifier constructions. The few studies available will be presented here.

Pizzuto (2002) addresses the acquisition of numerosity in a cross-sectional study of four Deaf children of Deaf parents aged 3;11 to 5;10, acquiring Italian Sign Language (LIS) using a picture description task to elicit specific LIS structures. The use of (1) inflectional and (2) uninflectional nouns and verbs in LIS was investigated. Inflectional nouns and verbs are defined as those where inflectional morphology is either optional or obligatory; uninflectional nouns and verbs are unable to accept inflectional morphology. LIS has three options to express numerosity: (1) adding the quantifier MANY (both inflectional and uninflectional nouns); (2) expressing numerosity by a change of location and movement, such as "repeating" the noun in sign space (inflectional nouns); and (3) expressing numerosity by the use of classifier constructions, such as CAR-ROW-ROW (inflectional nouns). All children in her study used the quantifier MANY added to the citation form of inflectional and uninflectional nouns to express numerosity. In the age span 3;11 to 5;10, they prefer to specify numerosity by adding an extra lexical marker rather than using a morphological device. Few developmental trends could be observed, although in the youngest child (3;11) the production of inflected nouns was entirely absent, while the other three children (5;5–5;10) produced some inflected nouns. However, it was not clear what these inflections referred to. In sum, this study suggests that there is a preference in children 4–6 years to use a quantifier to express numerosity in nouns over inflectional pattern in LIS. This study provided no information on distributional aspects, which would be represented in complex classifier constructions.

There are several studies that focus on the acquisition of classifier constructions, some also including the placement of multiple objects in space, such as CAR-ROW-ROW-ROW, but none of these studies deal separately with the acquisition of these constructions to express numerosity (e.g., Bernardino, 2005).

Fish et al. (2003) conducted a cross-sectional study on the acquisition of handshapes in classifier construction in ASL using the Real Object production task of the ASL Assessment Instrument (Hoffmeister, 1999) with 144 Deaf children of Deaf and hearing parents ages 3–12 years. Fish and her colleagues were looking specifically at classifier constructions in ASL that depict plurals and spatial arrangements of different objects (cars, pencils, and cans). No detailed analysis was provided, but the authors indicated that the younger children (3–5 years) achieved a score of 32% correct compared to 49% for the oldest children (9–12 years) across all items of the Real Object task (Fish et al., 2003). In sum, number and distribution undergo development, but no detailed information is provided.

Similar results have been found in a cross-sectional study by Hoffmeister (1992), indicating that the performance of Deaf children showed an age effect in receptive knowledge of plurals and arrangements in ASL using a comprehension task of the previously named ASL Assessment Instrument. Children aged 10–16 years performed better on this task than 8–9-year-old children. This result can be supplemented by a cross-sectional study by Hoffmeister, Philip, Costello, and Grass (1997) using different tasks of the ASL Assessment Instrument indicating that age was correlated with scores on a receptive and expressive measure of plurals and arrangement in ASL.

In sum, no detailed information on the acquisition of number and distribution is available. In Deaf children 4–6 years acquiring LIS there seem to be a preference to use the quantifier MANY over inflectional morphological devices (Pizzuto, 2002). Studies of ASL suggest that older children perform better than younger children in receptive and expressive tasks using plurals and arrangements (Fish et al., 2003; Hoffmeister, 1992; Hoffmeister et al., 1997). Studies of the acquisition of classifier constructions, in this case spatial verbs with whole entity classifiers, implicitly address the issue of distribution and numerosity, but do not address this issue separately (e.g., de Beuzeville, 2006; Schick, 1987).

In the following section, studies addressing negation in DGS and cross-linguistic differences will be presented.

#### 2.5.1.10 Negation in DGS and Cross-Linguistic Differences

Looking at the small body of research available for DGS, it is possible to find a few studies on negation (Pfau, 2001, 2004; Pfau & Quer, 2002, 2007). Pfau (2001, 2004) and Pfau and Quer (2002, 2007) state that negation in DGS uses a combination of manual and non-manual features (e.g., the manual sign NICHT (not) accompanied by a negative side-to-side headshake). The headshake often occurs simultaneously with a manual predicate. In most instances, the negative headshake is produced simultaneously with the verb, but it can also spread over an entire sentence. The manual sign NICHT follows the verb. While the headshake in a negated sentence is obligatory, the manual sign NICHT is optional. The combination of a headshake and negative manual sign occurs in many sign languages. In DGS, the negative headshake can accompany the verb on its own (in the absence of a negator sign) or co-occur with the verb and the negative manual sign. The headshake cannot occur only over the negative manual sign (Pfau & Quer, 2007).

Regarding the negation of modals in DGS, Pfau and Quer (2007) state that modals, for example, DARF (may) and WILL (want), cannot be negated by a headshake or a manual sign. Instead, they require a special negative form of the modal (a change in movement), accompanied by a headshake. Heßmann (2001a) has analyzed negation in a corpus of naturalistic DGS conversations. He observes that the side-to-side headshake in a DGS utterance is often combined with a negative facial expression. He also reports the use of several DGS negation signs, such as two variations of KEIN (none), three variations of NEIN (no) and NICHT (not), three variations of NICHTS (nothing), and one variation of NOCH-NICHT (not yet). Heßmann (2001a) observes that the negative headshake often occurs simultaneously with the part of the utterance being negated. However, it can also be produced simultaneously with negative manual signs like NICHTS. This has the function of emphasizing negation that has already been marked with a negation sign. Heßmann also reports that there is a group of signs that can be negated through change in movement, such as the sign STIMMT (right) can be negated by adding a twisting movement to the original downward movement to express STIMMT-NICHT (not right).

The different forms of negation – as presented in Heßmann's corpus – were analyzed for this study in regard to (1) their meaning, and to see if (2) they represented a dialect version or not. An attempt was made to match the BSL negation structures that are represented in the BSL test with comparable DGS structures, but because of the absence of cross-linguistic research, it is hard to tell if these are really equivalent forms. None of the negation signs and their variants were considered to be dialect forms (Heßmann, 2001a); variation is associated with differences in meaning. The studies by Pfau (2001, 2004), Pfau and Quer (2002, 2007), and Heßmann (2001a) were used as the basis for the adaptation into DGS of the BSL items expressing negation.

There is no cross-linguistic research on negation that directly compares BSL and DGS, but research comparing negation in a number of different sign languages is available (e.g., Pfau & Quer, 2002, 2007, 2010, in press; Zeshan, 2004, 2006). Zeshan (2004, 2006) undertook a typology study of negation (and other constructions) in 38 different sign languages.

Non-manual negation includes head movements and facial expressions that are suprasegemental (i.e., they spread over strings of manual signs of varying length). The non-manual markers in sign languages are facial expressions and head movements. Zeshan (2004) states that "in many cases, the form of these signals tends to be very similar across sign languages, whereas the status and scope of non-manual negative marking can differ quite radically" (p. 10). She reports three different types of head movement: (1) side-to-side movement of the head; (2) single sidewards head turn with the head remaining in this position (probably a reduced form of the side-to-side headshake); and (3) a backward tilt of the head, which has been found in the Eastern Mediterranean in Greek, Turkish, and Lebanese Sign Languages. The most common negative non-manual marker across sign languages is the side-to-side headshake. In addition to head movements, different facial expressions also occur in negative clauses. However, "their status as grammatical marker is often questionable…. Accordingly, they tend to be less obligatory and more variable than head movements" (Zeshan, 2004, p. 12). Although head movements occur with great regularity in negative clauses in the vast majority of sign languages, their grammatical status differs in different sign languages. For example, the side turn of the head in BSL only occurs with a specific negation sign. Results for a large number of sign languages indicate that a "negative head turn is dependent on a manual negative sign that it co-occurs with because it is too weak to function as a negator on its own" (Zeshan, 2004, p. 17), whereas side-to-side head movement and head tilt are independent negators. In the majority (26 of 38) of the sign languages in Zeshan's (2004) corpus, headshake-only negation across clauses is possible, but its frequency is quite variable. At one end of a continuum are sign languages that use headshake negation as the most frequent means to negate a clause, such as Norwegian and Swedish Sign Languages; at the other end of the continuum are sign languages where headshake-only negation is possible but relatively uncommon, such as Spanish Sign Language. "A combination of manual and non-manual negation is probably the most common strategy crosslinguistically, followed by headshake-only negation. Manual-only negation occurs rarely and is uncommon or impossible in several sign languages" (Zeshan, 2004, p. 18).

Besides the varying grammatical status of the headshake in different sign languages, the scope of side-to-side head movements differs across sign languages. The most common scope for headshake is either the whole clause or the clause without topicalized constituents (this feature occurs in 14 sign languages). In many sign languages, a headshake within a clause is not permitted. There are sign languages with more complicated scope rules. For example, in French-Canadian Sign Language, "the headshake cannot begin before the verb and cannot be stopped before the end of the clause" (Zeshan, 2004, p. 21). In DGS, the negator has a clause-final position, as in many other sign languages, but DGS "seems to be peculiar in this respect, since a headshake scope on the manual negator only is disallowed" (Zeshan, 2004, p. 24).

Also of interest is the relationship between manual and non-manual negators and how they are combined with each other. For example, in Greek Sign Language, "the backward head tilt preferably occurs with signs that involve an upward or backward movement of the hand" (Zeshan, 2004, p. 26); the rules seem to operate at a phonological level. Other sign languages have rules applying to individual negation signs. For example, in Ugandan Sign Language the headshake is unlikely to occur with the signs NOTHING and ZERO, although it occurs with other negator signs.

Sign languages use different kinds of negative particles:

> almost all sign languages in our data have a negative particle that conveys basic clause negation in our sense. In some cases, however, it seems as though non-manual negation should really be considered as the most frequent and/or the most unmarked way of negating a clause. (Zeshan, 2004, p. 30)

In French-Canadian Sign Language and the Scandinavian sign languages, manual negator signs are rare, but the headshake is frequent (Zeshan, 2004).

Zeshan (2006) provides an overview of the characteristics of sign languages that use the manual dominant or non-manual dominant system of negation. Sign languages that use the manual dominant system of negation preferably express negation with manual means, for example, by specific negator signs. In contrast, sign languages that use the non-manual dominant system express negation preferable by non-manual means (Table 2.5).

Table 2.5: Manual and Non-Manual Dominant Systems of Negation

| *Characteristics of non-manual dominant system of negation* | *Characteristics of manual dominant system of negation* |
|---|---|
| Non-manual negation is obligatory | Non-manual negation is not obligatory |
| Clause can be negated non-manually only, manual basic clause negator is optional | Clause cannot be negated non-manually only, manual negator is required |
| Choice of non-manual marking does not depend on manual signs | Choice of non-manual marking depends on choice of manual clause negator (if there is more than one non-manual negator) |
| Non-manual negation spreads freely over the clause | Scope of non-manual negation is over the manual negator only or is closely tied to the manual negator |
| Examples: | Examples: |
| German Sign Language<br>Swedish Sign Language<br>American Sign Language | Bali Sign Language<br>Turkish Sign Language<br>Japanese Sign Language |

Table from Zeshan (2006, p. 46), reprinted with permission from Ishara Press.

Across most sign languages, the negator is most often in a post-predicate or clause-final position (27 sign languages where data is available allow clause-final position of the negative particle). For some signs languages (e.g., Indo-Pakistani Sign Languages), this is the only acceptable position. Interestingly, in some sign languages (mostly European sign languages and the languages related to them in Australasia and the Americas, numbering 15 sign languages), the negator can also be in pre-predicate position.

> In sign languages with more than one position for the negative, the choice of position sometimes depends on which negative is involved. For example, Hong Kong Sign Language allows for the basic negator NOT to occur in pre-predicate position, while other negators are clause-final. (Zeshan, 2004, p. 40)

Other patterns have been noted by Zeshan (see also Heßmann, 2001a; Pfau & Quer, 2007) in DGS where some signs form negation by “modifying the movement to constitute a downward and diagonal inward-outward pattern” (Zeshan, 2004, p. 44).

Zeshan’s study is the first involving a large number of different sign languages and thus is highly relevant for test development and adaptation. It shows on the one hand, examples of how sign languages resemble or differ from each other, but it also shows on the other hand, how little is known about actual differences and similarities across sign languages. This study highlights the importance of knowing how different or similar the source and target languages are before adapting a sign language test.

#### 2.5.1.11 Acquisition of Negation

Reilly and Anderson (2002) discuss the acquisition of non-manual and manual morphology of negation in ASL by native signers. In ASL, the negative headshake across a negated predicate is obligatory; a negative manual sign can follow it but is not obligatory. ASL also has additional lexical markers to express negation, such as NOT and NONE (Anderson & Reilly, 1997). A negative manual sign must be accompanied by the negative non-manual headshake. Omission of the negative headshake with a predicate turns a negated sentence into a declarative sentence (Reilly & Anderson, 2002). There is also a small class of verbs in ASL that express negation by incorporation of a negative marker, such as KNOW / DON’T KNOW, WANT / DON'T WANT, or LIKE / DON'T LIKE (Anderson & Reilly, 1997). These resemble those signs described in DGS as taking negative incorporation (Heßmann, 2001a; Pfau & Quer, 2007).

The side-to-side headshake found in ASL is similar to gestural communicative headshakes in discourse. The difference is that a communicative negative headshake can occur independently, for example, as a response, and is thus independent of a linguistic utterance. Another difference is that the use of communicative headshake is inconsistent and inconstant regarding timing of onset and offset (Anderson & Reilly, 1997; Reilly, 2006). There are no specific rules when the communicative headshake is used, while the onset and offset of the grammatical headshake is very clear (i.e., it must always be coordinated with a single manual sign or utterance) (Anderson & Reilly, 1997; Reilly, 2006). The grammatical headshake always co-occurs with a manually articulated utterance, and is timed to co-occur with the predicate and a manual negative sign. The linguistic negative headshake is obligatory in sentences of negative polarity: (1) utterances with a manual negative sign; (2) utterances with an incorporated negative sign; and (3) predicates without manual negation (Anderson & Reilly, 1997; Reilly, 2006). As well as the headshake, configurations of the eyebrows, mouth, and nose convey negation. However, Anderson and Reilly (1997) conclude that the use of non-manual markers other than headshake are less clear-cut.

Anderson and Reilly (1997) conducted two studies to investigate the course of development of negation in Deaf native signing children. The first study used a cross-sectional design. Subjects were 51 Deaf children of Deaf parents aged 1;0 to 4;11. The children were filmed in a natural interaction (free play) with their mothers, but a sentence repetition task was also used. The communicative negative headshake was acquired by the age of 12 months, but did not co-occur with a manual negation sign. The first manual negation sign was produced between 18–20 months. These did not co-occur with non-manual features. At 19 months, the first verbs with incorporated negation emerged (e.g., DON'T WANT), but without headshake. Between 20–23 months, the first manual negation sign with headshake and the first verb with incorporated negation co-occurred with headshake. The first predicate with negation headshake was produced at 21 months, but the linguistic headshake was not correctly timed with the manual sign. It is only at the age of 27 months that children produce a predicate with a correctly timed headshake. There is an interval ranging from 1 to 8 months between the first occurrence of a manual negation sign and the co-occurrence of the sign with a headshake. In other words, children first acquire the communicative headshake, then manual negation signs, before they are able to integrate non-manual grammatical behavior with manual behavior. Reilly and Anderson (2002) state that the results suggest that "manual signs are acquired independently and grammatical facial behaviors appear subsequently as bound morphemes" (p. 169). This suggests a developmental process for non-manual features of "hands before faces" (Reilly & Anderson, 2002). In their second study, Anderson and Reilly (1997) used a longitudinal approach, filming 16 children at two different time points, using the same procedure as the first study. The results confirmed the findings of the first study.

The development of negation grows as the lexical and grammatical development of the children progress to increased use of manual negation signs. Later "when children acquire syntactic structures that include manual and non-manual signals, they invariably produce the manual component first, and only later do they add the non-manual behaviors" (Anderson & Reilly, 1997, p. 425). In other areas of syntactic acquisition, non-manual behaviors follow the acquisition of manual structures (Snitzler Reilly, McIntire, & Bellugi, 1991).

These results indicate the emergence, but not the mastery of negation in ASL. One aspect is the acquisition of manual negation signs, which go through different stages in the lexical development of Deaf children (Anderson & Reilly, 2002). The other aspect is the integration/co-occurrence of non-manual features (1) with manual negation signs, (2) in verbs with incorporated negation (DON'T LIKE), and (3) in sentences without manual negators.

Anderson and Reilly (1997) report:

> There is insufficient data to allow us to make strong statements about their acquisition of use. The performance of the children in their ability to integrate the linguistic headshake with the co-occurring manual sign was extremely variable. By the age of 4½, children were still making errors with respect to the timing of the manual and non-manual components. (p. 427)

The lexical development of the most common negation signs are acquired by the age of 35 months (Anderson & Reilly, 2002), but there is insufficient research on the acquisition of negation in complex predicates (Anderson & Reilly, 1997).

Both studies provide a good basis for the inclusion of the negation items that are represented in the adapted DGS test.

#### 2.5.1.12 Evidence for Other Structures Acquired in DGS

Regarding the development of DGS during the school years, only descriptive DGS progress reports of Deaf children from the bilingual pilot classroom from the School of the Deaf in Hamburg are available (Günther, 1999; Günther & Schäfke, 2004). These reports do not provide sufficient information on DGS development in areas of the DGS Receptive Skills Test to influence test adaptation. The individual progress reports cover areas such as communicative competence, narrative competence and the like, but in the absence of age-related developmental norms in DGS, information on the child's progress can only be provided in retrospect.

### 2.5.2 Summary of Cross-Linguistic Differences and Sign Language Acquisition

This review and analysis of the studies on sign language acquisition, in their relation to the linguistic structures represented in the adapted DGS test, has provided a first overview of the complete developmental path (i.e., from emergence to mastery) of some structures. This review has also identified not only the lack of sign language acquisition studies in general, but more importantly, the lack of cross-linguistic acquisition studies in comparison to the availability of studies comparing the acquisition of a sign language with different spoken languages (e.g., Morgan et al., 2006).

The studies on DGS suggest that many of the structures described in other sign languages are also available in DGS (e.g., classifier constructions), while others probably do not exist in DGS (noun/verb distinction: Becker, 2003). In turn, other structures in DGS have some features in common with BSL, such as verb agreement, but also have language-specific features (PAM). Other studies – not only addressing DGS – provide a good overview of cross-linguistic differences (Zeshan, 2004). These are important findings for the adaptation of the BSL test to DGS. The BSL structures and their occurrence in DGS are summarized in Table 2.6.

Table 2.6: BSL and DGS Structures in Comparison

| *BSL* | *DGS* | *Status* |
|---|---|---|
| Spatial verb morphology | Spatial verb morphology | Comparable structures available, but also language-specific differences (e.g., PAM in verb agreement) |
| Number and distribution | Number and distribution | Comparable structures available, but also language-specific differences (e.g., simple reduplication at same location of noun sign like HAUS++) |
| Negation | Negation | Comparable structures available, but also language-specific differences (e.g., change of movement in sign like KANN-NICHT to express negation) |
| SASSes | SASSes | Comparable structures available / "identical" |
| Handling classifiers | Handling classifiers | Comparable structures available / "identical" |
| Noun/verb distinction | | Exact linguistic status not determined in DGS |

### 2.5.3 The Role of Input on Sign Language Acquisition

The different linguistic experiences of Deaf children (e.g., Marschark, 2002) have been highlighted in Chapter 1 ("Introduction"). Only 5% of Deaf children have native signing parents (Mitchell & Karchmer, 2004), and might acquire a sign language as their L1, while Deaf children of hearing parents might have later access to a sign language and so undergo late L1 acquisition. In the light of Mayberry and her colleagues' (Mayberry, Lock, & Kazmi, 2002) finding that age of acquisition has an impact on the acquisition of ASL, it may be expected that the variable of age of exposure to DGS might result in different test performances in Deaf children with late vs. early L1 acquisition of DGS.

Different studies have investigated the effect of early vs. late L1 acquisition on language development and the long-lasting effect on language learning throughout the life of the Deaf individual. Mayberry et al. (2002) compared the language learning capacities of Deaf and hearing individuals as a function of early language acquisition. In two studies, Mayberry et al. (2002) address the question of whether early experience with a spoken language can facilitate subsequent learning of a sign language. Mayberry and her colleagues compared two groups of adults: (1) hearing from birth, who had learned English from an early age and started to learn ASL after becoming Deaf between 9 and 15 years old, and who had used ASL for more than 20 years; and (2) born Deaf, with little experience of ASL before they entered school and who had also had ASL experience for at least the last 20 years. The Deaf adults with little experience of language in early life showed low levels of ASL performance; the late-deafened adults, with English as a first language, showed high level of ASL performance. In sum, early L1 acquisition has a positive impact on later L2 acquisition, and delayed first language acquisition results in lower performance in ASL.

In the second study, Mayberry et al. (2002) investigated whether early experience with a sign language facilitates the subsequent learning of a spoken language. In this study, the following three groups were compared and all three consisted of adults who had learned English at comparable ages between 4–13 years and used it for over 12 years: (1) subjects were born Deaf and had little language experience before being exposed to ASL in school (late first language acquisition); (2) subjects were born Deaf born and had ASL experience since infancy (early first language acquisition); and (3) hearing adults with various spoken languages as their first language (early first language acquisition). The last two groups with language experience from early on showed higher levels of later learned English as L2. In contrast, the late learners of an L1 (ASL) showed lower levels of performance in English.

> [The results suggest] that the ability to learn a language arises from a synergy between early brain development and language experience, and is seriously compromised when language is not experienced during early life.… The timing of the initial language experience during human development strongly influences the capacity to learn a language throughout life.
> (Mayberry et al., 2002, p. 38)

Mayberry and Lock (2003) confirmed these findings in another study, which suggests that once the acquisition of a first language is in place, it is easier to acquire a second language. This is an important issue in the language development of Deaf children.

An important issue in early vs. late L1 acquisition is to determine the age span that accounts for early and late L1 acquisition. Plasticity of the brain in some domains gradually decreases and, for example, some domains of language such as phonology are more strongly affected than others, such as the lexicon (Fischer, 1998; Mayberry, 1993, 1995; Mayberry & Eichen, 1991; Newport, 2002; Newport, Bavelier, & Neville, 2001). Although plasticity of the brain is a process in which a peak period of plasticity occurs at "some maturationally defined time in development, followed by reduced plasticity later in life" (Newport et al., 2001, p. 482), an approximate age can be identified for this process: "with increasing age of exposure there is a decline in average proficiency, beginning as early as 4 to 6 and continuing until proficiency plateaus for adult learners" (Newport, 2002, p. 738). This age, between infancy and early childhood, is the critical time for successful (early) first language acquisition, with long-lasting effects on language performance in the L1 and later language learning, independent of the modalities of the first and second language.

So far, only the effect of delayed first language acquisition of a sign language has been reviewed. Goldin-Meadow and Mylander (1990) found that Deaf children of hearing parents developed a systematic gestural system as their means of communication. Even in a study across cultures (Goldin-Meadow & Mylander, 1998), the gestural system developed by Deaf children in the US and Taiwan shared a number of structural similarities and resembled to a certain degree natural language structure at sentence level. The hearing parents with Deaf children in two different cultures communicated with their Deaf children in speech, but a lot of interaction took place in action and gesture. The Deaf children in both countries "conveyed their message through gesture sentences rather than single gestures" (Goldin-Meadow & Mylander, 1998, p. 279). Spencer and Harris (2006), reviewing studies on gestural communication systems, concluded that "despite evidence that children have innate tendencies to construct functional communication systems from even somewhat degraded input …, there are clearly limits below which the input is insufficient to lead the development of well-articulated, formal language system" (p. 72). Singleton and Newport (2004) investigated the ASL development of a Deaf child with non-native signing Deaf parents who acquired ASL after age 15. The language learning of this Deaf child is based on the input from his late-learner parents. At the age of 7, this Deaf child performed better on an ASL morphology task than his parents, indicating that he is – despite inconsistent ASL input – able to acquire most ASL morphemes comparable to native signing Deaf children (for a detailed review, see Singleton & Newport, 2004).

In relation to the present study, the issue of early vs. late exposure to DGS in L1 has an impact on the adaptation of a sign language test. First of all, the variable age of exposure to DGS might have an impact on differences in performance on the adapted DGS test, and thus might contribute to explain a possible difference in the test performance of Deaf children who acquired DGS as late L1 as compared to their same-age peers, who acquired DGS in the process of early L1 acquisition.

Another question relating the studies of early vs. late L1 acquisition to test adaptation can be raised in relation to what constitutes the ideal norming sample for the adapted DGS test for this population. It may be the case that different norming samples are most appropriate depending on early and late acquisition of DGS as a L1. This will be further investigated in the last chapter ("Discussion").

Having provided an overview on the various acquisition and cross-linguistic studies and the issue of input on language acquisition, in the next section these various studies will be linked to sign language test adaptation.

## 2.6 Sign Language Acquisition and Test Adaptation

In the previous sections on studies of sign language acquisition, the linguistic structures that are represented in the BSL test and their comparable structures in DGS, as well as differences and similarities in these structures across sign languages, were discussed.

In this section, these topics will be discussed in relation to the adaptation of the BSL test to DGS. Justifications for test adaptation in general will be presented and the current state of research on sign language acquisition in DGS will be addressed, followed by a description of the state of research on the structures of DGS and differences across languages. It will be argued that despite the current limited amount of DGS research – a situation shared by many sign languages – research from better-documented sign languages can be used to build hypotheses for the adaptation of sign language tests. Finally, it will be shown how this approach circumvents the approach taken by Herman et al. (1999) in the development of the BSL Receptive Skills Test.

### 2.6.1 Reasons for Test Adaptation

There are two main reasons for adapting psychological tests or spoken language and sign language tests: (1) to conduct cross-cultural and cross-linguistic research; or (2) for economic reasons, such as to save money (Hambleton & Kanjee, 1995). A third reason is a pragmatic one, which specifically applies to the adaptation of sign language tests and the state of research on the sign languages in question. De Beuzeville (2004, 2006) points out a temporal relationship between the conducting of sign language research and the development of sign language tests. In a first step, research is undertaken on the structures of the language as adults use them. In a second step, research is done on how children acquire the sign language, and finally, sign language tests are developed based on the prior two steps.

De Beuzeville recognizes, however, the practical need to develop tests before steps one and two are completed, and argues that this is acceptable if the developer is aware of the issues concerning the validity and reliability of such instruments. Basically, the researcher undertaking the present study is in agreement with de Beuzeville's proposal, while also being aware of the urgent need in Deaf education for sign language tests in different countries (e.g., UK: Herman, 1998; Germany: Haug & Hintermair, 2003) despite the absence of descriptions of adult usage or acquisition studies. Therefore, it can be argued that test adaptation can be appropriately motivated by pragmatic reasons relating to the state of research in this field. Test adaptation in this situation does not involve a simple one-to-one translation from the source language to the target language, a process that has also been found not to work successfully in the translation of spoken language tests (Alant & Beukes, 1986; Chavez, 1982; Rosenbluth, 1976; Simon & Joinier, 1976).

### 2.6.2 Sign Language Acquisition Studies as the Basis for Test Adaptation

The literature on the adaptation of spoken language tests indicates that during the adaptation process, the results of both acquisition studies and studies of the adult language (plus a panel of experts) are used to make informed decisions about the following questions:

(1) Which items of source tests have a linguistic (and cultural) equivalent in the target language (will most likely be "simple" lexical items)?
(2) Which items need to be deleted because they represent a language-specific structure of the source language that does not exist in the "same" way in the target language?
(3) Which items need to be replaced by items that represent culturally-appropriate concepts in the target culture?
(4) Which items that represent language-specific structures of the target language that do not occur in the source language need to be added to the test (Friend & Keplinger, 2008; Hamilton et al., 2000; Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir & Ellis Weismer, 1996)?

These are important criteria and underpin the procedures required for the adaptation of a language test. As pointed out already, the situation for the adaptation of sign language tests is different and more complicated than the adaptation of spoken language tests, since more research is available for (most) spoken languages than for most sign languages. Results on the adaptation of a sign language test to another sign language clearly highlight two main issues that are relevant here: (1) language-specific issues, and (2) culturally related issues (Haug & Mann, 2008). In addition, many sign language tests are still under development and have not reported strong psychometric properties (Haug, 2008a).

As has been pointed out earlier, the existing research studies on the acquisition of DGS do not provide a sufficient basis for adapting a sign language test from the source language BSL to the target language DGS. In order to be able to approach the adaptation of sign language tests, the following dilemma needs to be resolved: On one hand, there is limited research on the acquisition of DGS available in order to make a clear statement about the emergence and mastery of DGS, which is the basis for test development/adaptation, and on the other hand, there are scientific requirements (e.g., reliability, validity) that need to be met in order to produce a test that can be used.

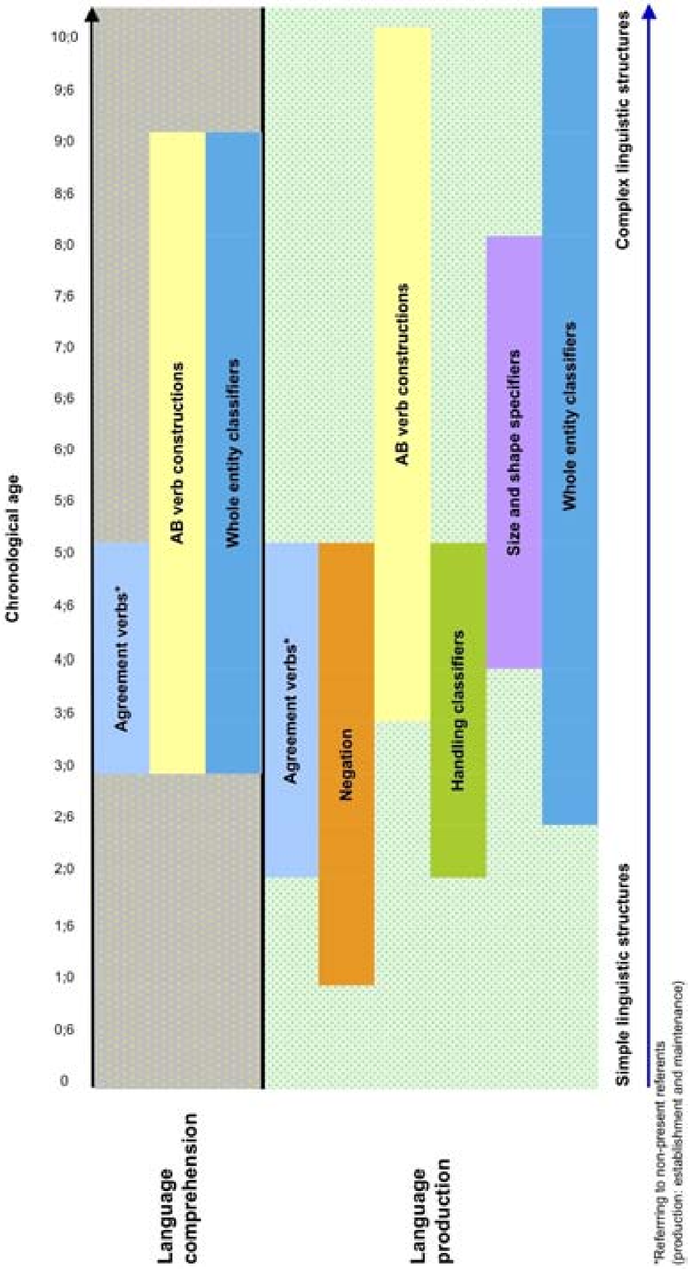
An overview of the most recent cross-linguistic literature on language acquisition (see Figure 2.1) illustrates the emergence, development, and mastery of some of the structures that are represented in the BSL test, structures which also should be represented in the adapted DGS test. It must be kept in mind that most of the studies reviewed for this overview do not refer to the acquisition of DGS (with the exception of Hänel, 2003, 2005), and while they provide initial evidence concerning the developmental pattern of sign language acquisition in ASL and BSL, they cannot be used to make a direct interference to DGS acquisition.

Another important issue is that most of the acquisition studies reviewed focus on language production, whereas the adapted DGS test evaluates language comprehension. Only a few studies also look at sign language comprehension (see Figure 2.1). For example, studies on the acquisition of the complex AB verb construction in BSL (Morgan & Woll, 2002b, 2003) looked both at comprehension and production data. The results revealed that comprehension precedes production, which is not surprising as this has also been found in studies of language comprehension and production in spoken languages (e.g., Hirsh-Pasek & Golinkoff, 1996).

Children acquire linguistic structures across certain time spans. It is known from child (spoken) language research that simple linguistic structures are acquired before more complex structures (Clark, 2003; Menyuk, 1988). Related to sign languages, for example, negation is initially manifested as a communicative side-to-side headshake before the first manual negator is produced; only later do Deaf children produce both channels simultaneously. The goal of the adapted DGS test is to evaluate exactly this *developmental aspect of different linguistic structures*. If certain structures reflect the language development of Deaf children ages 4–8, then it is important that the items of the test which represent these structures should differentiate between younger and older children. In other words, the test should represent linguistic structures that differentiate by age, for example, between simpler and more complex structures[21]. The reviewed studies on sign language acquisition thus provide the basis for informed decisions about which structures should be represented in the adaptation of the test item from BSL to DGS: that is, *to provide a basis that accounts for developmental aspects in the adapted DGS test*, but *not* as a baseline for DGS acquisition.

[21] The general goal of such norm-referenced tests is to see if a tested child performs comparably to his/her peers.

Figure 2.1: Overview of Sign Language Acquisition of Reviewed Studies



In summary, the reviewed studies on sign language acquisition suggest an approximate developmental timetable for structures, from emergence to mastery, that are represented in the BSL test which can be carried over to the adaptation to DGS. Even though most of these studies provide information only on production, it can be assumed that comprehension precedes production. Therefore, it is argued that the findings of developmental aspects drawn from other studies on sign language acquisition can be used as a basis for informed decisions on what should be represented in the adapted DGS test.

### 2.6.3 Cross-Linguistic Differences and Similarities

Another relevant issue concerns *similarities and difference across sign languages* as found in adult language. It was argued that the reviewed studies of sign language acquisition could be used to account for developmental features in an adapted test in order to differentiate between younger and older children. At this point it is also important to look at the similarities and differences across sign languages and how this feeds into the adaptation of sign language tests.

There are some studies that, for example, compare specific aspects across two or three sign and spoken languages, such as negation (e.g., Pfau & Quer, 2002, 2007), verb agreement (e.g., Mathur & Rathmann, 2001; Rathmann, 2003; Rathmann & Mathur, 2002), or sign language acquisition in comparison with spoken language acquisition (e.g., Morgan et al., 2006). A large typological study on negation and interrogatives constructions in 38 different sign languages (Zeshan, 2006) has also been undertaken. The cross-linguistic studies reviewed here do not compare BSL and DGS directly, but provide good examples of the differences between sign languages (Zeshan, 2006).

Studies that address a specific structure in BSL and studies that address the same structure in DGS suggest that there are both common features and language-specific features. An example of this is the sign glossed as PAM in DGS (Mathur & Rathmann, 2001; Rathmann, 2003; Rathmann & Mathur, 2002), which does not have a counterpart in BSL. Such a language-specific feature needs to be represented in an adapted DGS test.

There are also structures across sign languages that have few language-specific features, in particular the representation of objects and events in space utilizing classifier constructions (e.g., Sandler & Lillo-Martin, 2006; Schembri, 2003). It is not argued here that classifier constructions are *alike* across sign languages, but in comparison to other more language-specific structures, these features are similar across sign languages. This similarity is probably motivated by the iconic representation of object shapes or how they are handled.

The main point here is that although cross-linguistic research in sign language acquisition is still in its infancy, for some of the structures in the test, research is available that provides insight on the differences between BSL and DGS, as well as similarities (e.g., handling classifiers). These studies suggest that some of the differences and similarities between BSL and DGS might also be applied more generally across sign languages.

### 2.6.4 Building Hypotheses

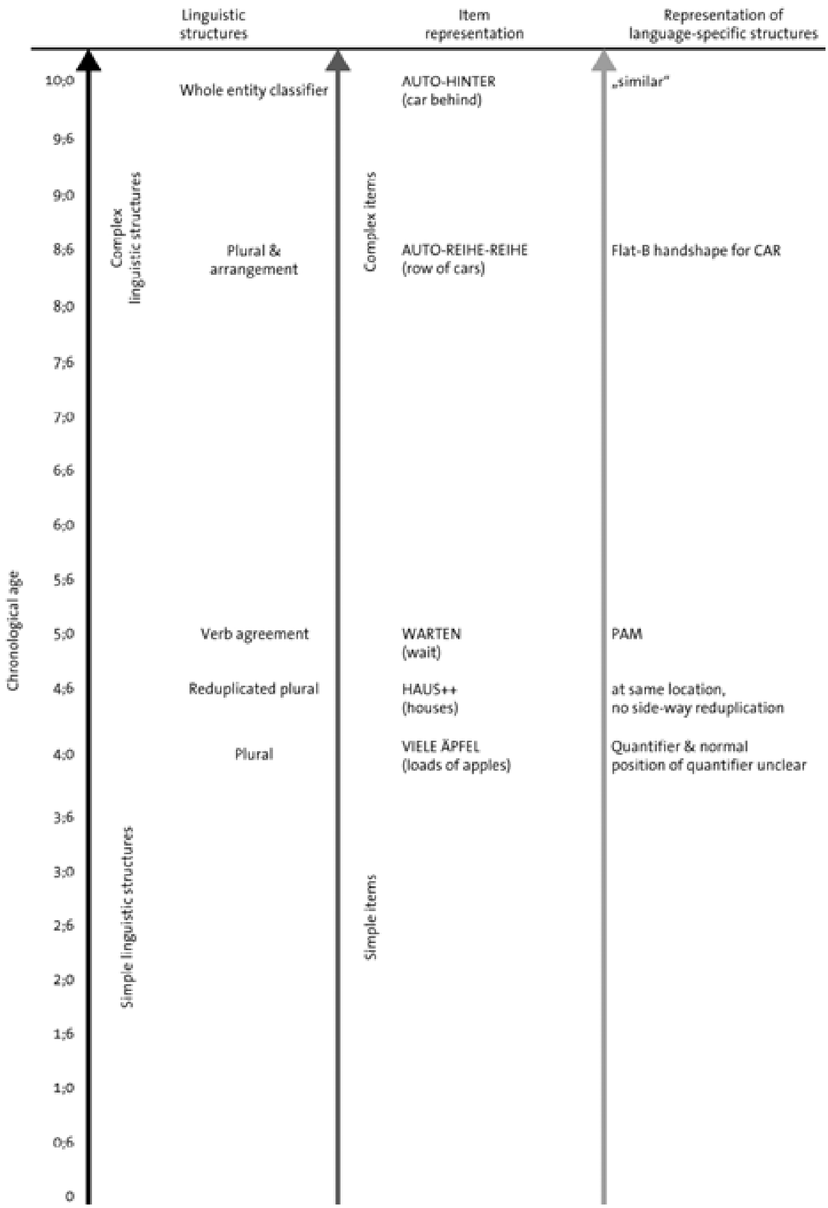
Based on the reviewed (1) acquisition studies, (2) cross-linguistic studies, and (3) DGS studies, a hypothesis-building approach for the adaptation of sign language tests will be proposed (Figure 2.2). The findings which have been discussed are summarized below and form the basis for the hypothesis.

#### Hypotheses

(1) *Cross-linguistic studies of sign language acquisition indicate that simpler linguistic structures are acquired at an earlier age than are complex linguistic structures and thus reflects language development.*
(2) *Simple and complex linguistic structures are represented by simple and more complex items in the target test, which in turn reflect language development.*
(3) *Based on the research review, language-specific structures are represented in these target items.*

The next step is to operationalize these hypotheses and use them as the basis for deciding which items should be represented in the adapted DGS test. These hypotheses can also be illustrated visually (see Figure 2.2: Map of Ranking of Item Complexity). These formulated hypotheses will be integrated as a basis for test adaptation into the methodological part and will be examined/verified implicitly for the most part, but also, in part, explicitly.

Figure 2.2: Map of Ranking of Item Complexity



### 2.6.5 The Different Approaches Taken by the BSL Test and the Adapted DGS Test

The most obvious difference between the development of the BSL Receptive Skills Test and the adapted DGS test is that the former is a new development, whereas the latter is an adaptation of the former. Still, it is important to have a closer look at the different methodological approaches taken in the development and adaptation of these tests.

For the development of the BSL test items, studies on the acquisition of BSL and ASL were reviewed in order to identify linguistic features that are important for the acquisition of BSL. The authors justified using ASL acquisition studies as a basis for developing items for the BSL test because of the similarity between these two languages (Herman, 2002). The development of the BSL items is based on the *notion of similarities* between these two sign languages, which in turn provides empirical evidence for developing test items. The reliability and validity of these test items were confirmed in the process of the test development (Herman, 2002).

Adapting the test from BSL to DGS also has the consequence that certain facts are already specified, as, for example, the targeted linguistic structures. However, adapting BSL test items representing some BSL structures, such as negation, requires a careful review of these structures (and acquisition) in DGS. There is no study that clearly documents the differences and similarities between BSL and DGS. Therefore, a hypothesis-building approach is used here for the adaptation of the BSL Receptive Skills Test to DGS. A major difference to the approach applied in the development of the BSL test is that for the adaptation of the DGS test, hypotheses are formulated as the basis for test adaptation. These hypotheses are based on (1) acquisition studies of other sign languages to account for the developmental aspect, and (2) DGS studies and cross-linguistic studies.

## 2.7 Summary and Implications for the Present Study

In the previous sections, important key concepts in language testing, as well as steps and procedures in the development and adaptation of language tests and how they relate to this current study, have been reviewed. Most steps and procedures, such as the chosen purpose or method, apply to the adaptation of the BSL test to DGS. Reviews of the adaptation of spoken language tests have provided an important basis for identifying potential sources of errors in the adaptation of a sign language test, and have been supplemented by a review of the adaptation of sign language tests to another sign language. An overview of available sign language tests provided a basis for defining criteria that led to the decision to adapt the BSL Receptive Skills Test.

This was followed by an overview of acquisition studies across sign languages that have addressed the structures represented in the BSL test, which are candidates for the adapted DGS test. These cross-linguistic studies also provide a general overview of the developmental path of the emergence and mastery of some of these structures. Studies of DGS structures have shown where similarities and differences are found. Studies of late first language acquisition shed light on the long-lasting consequences of late acquisition of sign language as an L1 on Deaf people's first and later second language learning competencies. These studies underscore the importance of early access to a sign language and how this might explain performance differences in Deaf children on the adapted DGS test.

Lastly, all the issues reviewed and discussed have been linked to the topic of adapting sign language tests, where it has been argued that on the basis of cross-sign language acquisition studies, studies comparing different sign languages, and studies of DGS structures, hypotheses can be formulated that can serve as the basis for test item adaptation.

This literature review has the following implications for the present study's research questions:

(1) Basic terms and concepts provide insights as to how to proceed in test adaptation. Studies of spoken test adaptation are useful in raising awareness of potential sources of errors in test adaptation. The spoken language test literature also clearly highlights the differences between the current state of research in spoken languages and sign languages. Studies of sign language acquisition together with cross-linguistic studies provide an overview of selected aspects that are represented in the BSL test and that should be adapted to DGS. The DGS studies are also an important element in the adaptation of test items from BSL to DGS. In sum, all of these different kinds of studies have implications for the adaptation process, and will also serve as the basis for analyses of the psychometric properties of this test.

(2) The review of acquisition studies and the issue of late sign language acquisition are especially relevant to the second set of research questions which address the impact of different variables on test performance. These studies served as a basis for the research questions concerning the extent to which age of exposure has an impact on the acquisition of a language and consequently also on test performance.

Having reviewed the literature relating to the present study, the next chapter will turn to the methodology and research design applied in this study.

# 3 Methodology

In this chapter, the methodology and the research design used in this study will be presented. The research design was an experimental approach using cross-sectional testing methodology (comparing children of different ages to each other at one point in time) and includes the test adaptation, that is, the test adaptation process from the source sign language, British Sign Language (BSL), to the target sign language, German Sign Language (DGS).

The purpose of this study was to address linguistic, cultural, methodological, and theoretical issues during the test adaptation process from BSL to DGS, which included several stages of piloting and revising the first DGS test version before the main study was conducted. Since this study was undertaken independently for a doctoral thesis, no resources were available to standardize the adapted DGS test.

This chapter is divided into five sections: (1) the instrument and Pilot 1 to establish the suitability of the test items; (2) Pilot 2 testing and revisions of the first version of the instrument; (3) description of the test items of the main study; (4) the main study, including the procedure; and (5) data analysis. In addition to providing information on the research participants, on the documentation of the adapted and revised instrument, and on the procedure to collect the data, the section on data analysis also contains information on the statistical procedures that will be applied in the "Results" chapter.

## 3.1 The Instrument

In this section, the instrument that was used for this study will be presented. The first step was Pilot 1, which was conducted with Deaf adults and children in order to check the suitability of the test materials. In the second step, the testing interface was programmed (first test version), and then Pilot 2 was conducted with (1) non-signing hearing children, and (2) Deaf adults, which lead to some revisions of the adapted test.

The decision for adapting an existing test rather than developing a new test was largely based on the fact that the research on acquisition of DGS is very limited. After a careful evaluation of available sign language tests (Chapter 2, Section 2.4) – it was decided to use the BSL Receptive Skills Test (Herman et al., 1999) as a template for the DGS Receptive Skills Test. The BSL test has already been presented in "Literature Review" (see Chapter 2, Section 2.4.5). The adaptation of the BSL test was only made possible with the permission of the authors. The BSL Receptive Skills Test (1) covers the age range 3–11 years old and therefore the appropriate linguistic structures, (2) it tests receptive skills, and (3) is a standardized test. Table 3.1 provides an overview on the test adaptation process.

Table 3.1: Overview of Test Adaptation Process

| *Steps* | *Description of steps* |
|---|---|
| 1. Review and revision of test stimuli | Picture materials were reviewed and changes were made (e.g., replacing the red British mailbox with a yellow German mailbox) |
| 2. Pilot 1 | Establish suitability of test items: Check for regional variation in three regions with Deaf adults and children |
| 3. Adaptation of items | (1) Order of test items<br>(2) Comparability of BSL and DGS linguistic structures<br>(3) Development of 10 additional items |
| 4. Filming of test | Filming of test instructions and test items |
| 5. Programming test interface | Programming of a user-friendly test interface that runs on a laptop and can store the results automatically |
| 6. Pilot 2 | Piloting first test version with:<br>(1) Non-signing hearing children and<br>(2) Deaf adults |
| 7. Revisions of first version | Revisions of the first version based on Pilot 2:<br>(1) Changes to the pictures<br>(2) Re-filming of items<br>(3) Changes to the layout |
| 8. Planning of main study | (1) Contacting the schools<br>(2) Development and distribution of educational back ground questionnaires for children |
| 9. Main study | Conducting the main study at five school sites in Germany |

### 3.1.1 Review and Revision of the Test Stimuli

The picture materials – for both the vocabulary check and for the receptive skills test of the template – were reviewed in order to see if any changes needed to be made for cultural reasons (e.g., Haug & Mann, 2008).

Deaf and hearing experts with a strong background in sign linguistics looked at the pictures and gave input on what needed to be changed. Changes fell into two categories: (1) cultural-related issues; for example, the steering wheel of a British car needed to be moved from the right to the left side of the car, or a British round red mailbox needed to be replaced by a square yellow German mailbox (Figure 3.1); and (2) concept representation, because what the picture was intended to depict was not clearly presented in the picture; for example, a queue of people was represented by only three people in the original version of the test (Item 24), and more people were added in the DGS version in order to better express the concept of *queue* (Appendix A-1).

Figure 3.1: Example of the Target Picture Used in the BSL Test (Left) and the Revised Picture Used in the Adapted DGS Test (Right)



A Deaf illustrator, who worked at Sign Language-Media in Zurich producing sign language teaching materials, updated the pictures accordingly.

### 3.1.2 Pilot 1 to Establish Suitability of Test Items

After reviewing the original test materials, it was decided to conduct a first pilot (Pilot 1) with those materials. All test materials from the vocabulary check, as well as the pictures for the receptive skills test, were shown to Deaf children and adults. The data were collected in three of the five regions where testing in the schools for the Deaf later took place.

The objectives of Pilot 1 were (1) to check for regional variations of lexical items for DGS (vocabulary items), and (2) to see how well the distractors would work for the adapted DGS version (receptive skills test items). Regional variations of signs need to be considered carefully when further developing the items since variation may require the creation of more than one version of the test for different regions (see Chapter 3, Section 3.1.9).

In order to cover as wide a range as possible of regional variation, the data for Pilot 1 were collected between September 2004 and February 2005 in three sites in Germany.

#### 3.1.2.1 The Testing Sites for Pilot 1

The choice of the three sites was based on the regional distribution of schools for the Deaf where data collection for the main study later took place. Full data were not obtained at all three testing sites. Table 3.2 provides an overview of the three sites and the number of participants at each site.

Table 3.2: Pilot 1 – Data Collection Sites and Materials

| *Region* | *Site* | *Vocabulary items* | *Receptive skills items* | *Participants (N = 13)* |
|---|---|---|---|---|
| Northern | Site #1 | yes | no | 3 (adults only) |
| Southern | Site #2 | yes | yes | 4 (adults only) |
| South-West | Site #3 | yes | yes | 6 (5 children, 1 adult) |

The age range of the 13 participants was 12 to 57 years ($M$ = 31; 4 male, 9 female).

#### 3.1.2.2 Procedure of Pilot 1

The data collection procedure was designed and piloted in collaboration with two Deaf sign language instructors. After the revisions, all interviewers received written instructions on the data collection procedure, which comprised the following parts:

(1) An explanation of the objectives of Pilot 1
(2) A description of the data collection procedure
(3) A questionnaire for obtaining background information about the participants (Appendix B-1)

All interviewers and participants (or their parents/legal guardian), signed a consent form (Appendix C-1 and C-2).

The interviewer presented the pictures of the vocabulary check and the receptive skills test on a laptop computer, and the participants were asked to sign what they saw. The interviewer and participant were videotaped during this session.

The three interviewers involved in Pilot 1 were Deaf and had a background in teaching German Sign Language. All data were edited and inserted into a specially designed File Maker data bank.

#### 3.1.2.3 Results of Pilot 1

Most of the items from the vocabulary check, which depict simple nouns, showed no regional variations. There was some variation in the signs JUNGE[22] (boy), KIND (child), HUND (dog), MUTTER (mother), and TEDDYBÄR (teddy), but these variants were not used consistently across informants in a single region, that is, the variations could not be clearly ascribed to one particular region.

All vocabulary items were discussed with two Deaf sign language instructors who evaluated their status as conventional lexical forms (Appendix D-1). A conventional lexical form in the present study was defined as a sign that has a consistent form-meaning relationship and which the Deaf sign language instructors know is used by adult signers in the Deaf community. Signs that were considered as not meeting these criteria included the following: (1) signs where a different but semantically related

[22] DGS glosses will be written in German with the English “equivalence” in parentheses in order to mark a distinction to glosses in BSL (or other sign languages).

manual form was produced with the mouthing pattern of the DGS target; for example, the DGS sign JUNG (young) was produced with the mouthing of the German word *Junge* (boy); and (2) signs considered to be home signs[23] which do not bear a conventionalized form-meaning relationship and are not used in the wider adult signing Deaf community[24]. The items that showed variation are presented in Table 3.3 (for an overview of all vocabulary items, see Appendix D-1). Summarizing these results, five lexical signs showed regional variation. Examples for the four different variants of the sign JUNGE (boy) are shown in the Figure 3.2 (for the other variants, see Appendix D-2).

The production data collected from the receptive skills test were not used in the test adaptation process because of the large variability. For example, younger informants tended to provide very brief descriptions of the pictures while adults often created whole stories out of the pictures. Because of these issues, it was decided not to use these data in the adaptation. As with other sign languages, it is assumed that there is greater lexical variation than morphological and syntactic variation (see Woll, 1998 for comments on variation in BSL). Some pictures (from the vocabulary check) triggered a different sign than the one expected, even when it was semantically related; for example, MÄDCHEN (girl) was signed instead of MUTTER (mother). This factor could not be controlled for.

23 The concept of *home signs* refers to a gestural communication system used by Deaf children of non-signing hearing parents with their families. The use of it can range "from simple pointing at objects and acting out messages, to a repertoire of agreed-upon gestures that convey a much more extensive range of information, sometimes even affective information" (Lane, Hoffmeister, & Bahan, 1996, p. 39). In the questionnaires used in the present study to investigate the means of communication at home and in school, the German terms *eigene Gebärden / Gesten* (own signs / gestures) were used. These German concept have been translated into English here as *home signs*.

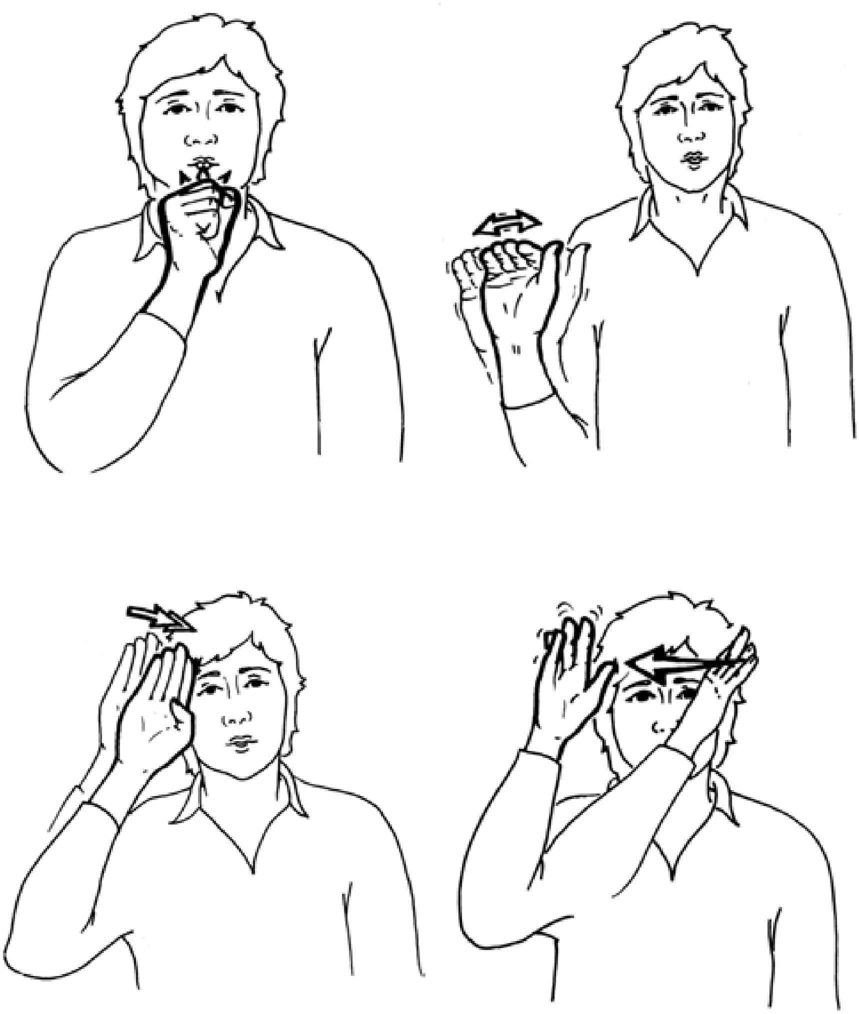
24 These results are only based on the analysis with two Deaf sign language instructors. The results are not representative in this sense, but suggest a first insight into lexical variations in DGS. With the ongoing 15-year DGS Corpus-Lexicon Project at Hamburg University that started on January 1, 2009, it will be easier in the future to obtain empirical data on regional variations. This data was not available at the time of the present study (for the DGS Corpus-Lexicon Project at Hamburg University, see http://www.sign-lang.uni-hamburg.de/dgs-korpus/homee.html, retrieved on April 20, 2009).

Table 3.3: Regional Variations and Conventionalized Forms of Vocabulary Items

| *Name of item / Variant* | *Vocabulary item** | *Results of analysis* |
|---|---|---|
| JUNGE1, 3, 4, 5 (boy) | 6 | JUNGE1: 6 informants (South-West)<br>JUNGE3: 7 informants<br>JUNGE4: 3 informants (North)<br>JUNGE5: 2 informants (South)<br>Four distinct lexical signs were identified as conventionalized forms, one variant could not be assigned to a certain region. |
| KIND1, 2 (child) | 7 | KIND1: 4 informants<br>KIND2: 2 informants<br>Both variants were considered as conventionalized forms, but could not be assigned to a specific region (South, South-West). |
| HUND1, 2, 4 (dog) | 11 | HUND1: 8 informants<br>HUND2: 3 informants<br>HUND4: 3 informants (South)<br>Six different variants were collected, but only three variants were considered as conventionalized forms. Two of these three variants could not be assigned to a specific region. |
| MUTTER1/MAMA (mother/mama) | 18 | MUTTER1: 3 informants (South)<br>(This stimulus elicited the signs for MAMA, FRAU (woman), and MÄDCHEN (girl), the majority of the informants signed FRAU, but the sign FRAU did not occur in the test). The sign MUTTER1 was considered as a conventionalized form and could be assigned to a specific region (South). |
| TEDDYBÄR3, 4, 5 (teddy) | 21 | TEDDYBÄR3: 2 informants (North)<br>TEDDYBÄR4: 1 informant (North)<br>TEDDYBÄR5: 4 informants (South)<br>Six variants were collected, but only variants 3, 4, and 5 were considered as conventionalized forms (in some cases the informants signed BÄR (bear), and not the expected sign TEDDYBÄR). |

* Items are numbered (e.g., JUNGE1 (boy) as they occur in the DGS Receptive Skills Test data bank); a different number is assigned to each variant. All items are glossed in German in the data bank.

Figure 3.2: Regional Variations of the Sign JUNGE (Boy): JUNGE1 (Top Left), JUNGE3 (Top Right), JUNGE4 (Lower Left), JUNGE5 (Lower Right)



In this section, Pilot 1, which was the basis for the adaptation to DGS of the BSL Receptive Skills Test and which was used to create the items for Pilot 2 and the main study, has been described. In the next sections, the item design of the first adapted version used in Pilot 2, is described.

### 3.1.3 Item Design

The instrument that was used in this study is an adaptation of the BSL Receptive Skills Test (Herman et al., 1999) to German Sign Language. However, the original BSL test is video-based, whereas the DGS Receptive Skills Test is computer-based. The first version of the adapted DGS test consisted of 22 items (as in the original BSL test) for the vocabulary check, and 53 items (including 3 practice items) in the receptive skills test. The original 40 items were adapted and 10 new items were developed in close collaboration with an advisory panel of Deaf and hearing experts. These items were integrated in the computer interface of the test. One item (Item 36, HEARING-AID NOTHING) had been already changed at this point. The hearing-aid was replaced by the sign for BALL. As a consequence, the vocabulary item for hearing-aid was also removed from the vocabulary check, leaving only 21 vocabulary items compared to 22 items in the BSL test.

Since the purpose of the original BSL test was to test the language development of Deaf children from 3–11 years old, the test needed to include items that differentiate between younger and older children in test performance. For example, the test needed to include items that are correctly responded to by younger and older children, but also items that are only correctly responded to by the older children, in order to evaluate language development over the age range for the test (cross-sectional). The 40 items of the BSL test reflect different levels of difficulty, and appear in order of their level of difficulty. The easiest ones are followed by the more difficult ones across different areas of BSL morphology and syntax. The 10 additional items were developed because some of the items in the original BSL test do not work in the same way for DGS (i.e., order of acquisition is different) and because some language-specific structures in the BSL test do not occur in DGS. Similar findings have been reported for the adaptation of the BSL Receptive Skills Test to Auslan (Johnston, 2004).

The 10 additional items were equally distributed across all linguistic structures of the BSL test. During the development of the 10 additional items, the main goal was to cover the same linguistic structures (negation, spatial verb morphology) covered in the BSL test. Thus the new items paralleled those in the original test (e.g., items on spatial morphology and with varying levels of difficulty). The development of those 10 additional items was done in close cooperation with Deaf experts. New pictures were created based on the style of the existing pictures. The order of the original items in this first adapted DGS version followed the order of the items in the BSL test, followed by 10 newly developed items. The 10 additional items were added at the end, since it was unclear how their level of difficulty would fit into the overall order of difficulty.

In the final version at this stage, (1) three additional items covered spatial verb morphology, (2) two items were concerned with size and shape specifiers, (3) three items with number and distribution, (4) one item with negation, and (5) one item with a noun-verb distinction (for a complete list of all 53 items, see Appendix D-3).

### 3.1.4 Item Format

In contrast to the video-based testing format of the BSL Receptive Skills Test, which included the coding of the subject's responses on a scoring sheet, it was decided to use a computer-based testing format based on the following reasons. First, this approach makes it possible to save the test results automatically, which makes the test more time efficient and minimizes the number of errors during the data entry. Secondly, the standardized format makes it possible to minimize effects of variation in testing conditions since all test instructions as well as test items are included in thecomputer-based DGS test.

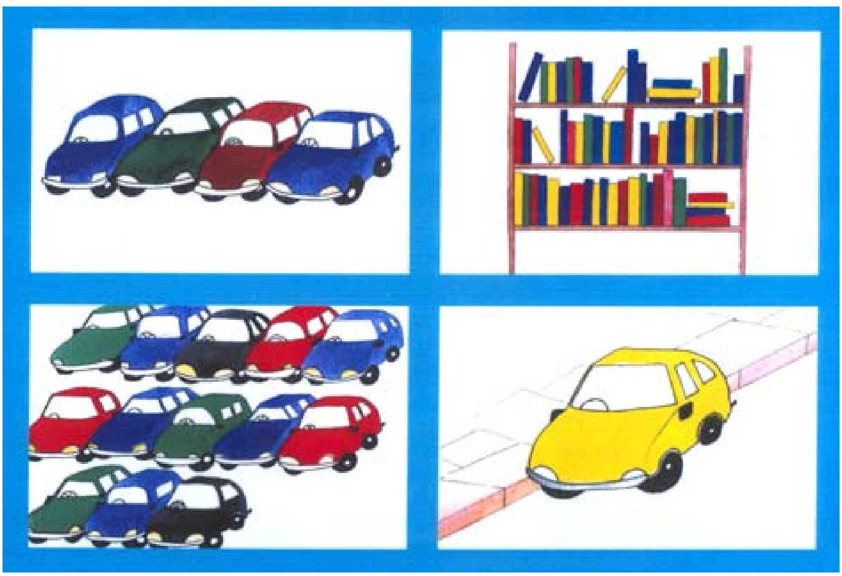
The vocabulary check was based on that in the BSL test version, but delivered on the computer.

The item format was a fixed response format, that is, a multiple-choice format. This item format was determined by the test template where each item consisted of three or four possible answers that could be chosen by the child. In order to minimize the chance that a child could reach the correct answer by chance, alternative answers (distractors) were used throughout the test. This design requires the subject to select the correct response from two to three alternatives that are phonologically, lexically, or morphologically related to the correct answer.

### 3.1.5 Types of Distractors

Alternative answers, or distractors, needed to be provided for each item in order to minimize the selection of correct answers by chance. The multiple-choice format means that one out of three or four pictures is correct, and the other choices are incorrect. The wrong answers need to be similar enough to the target that selection of the target is not overly simple, but not so similar that it is difficult to identify the target. Different types of distractors were used. An example is given below for the signed stimulus in German Sign Language AUTO-REIHE-REIHE-REIHE (*cars parked in rows)* (Figure 3.3).

Figure 3.3: Example for Distractors of the BSL Receptive Skills Test



© Herman et al., 1999

The right answer (lower left) shows *three rows of cars*. The image on the upper right depicts *books on shelves*, and is a phonological distractor because the hand configuration is the same as in the target *cars parked in a row*, yet the orientation of the hand is different from that in the target. The pictures in the upper left and lower right are morphological distractors because they differ from the right answer in number and orientation in space.

### 3.1.6 Item Representation

In the previous chapter ("Literature Review"), the BSL Receptive Skills Test, the linguistic structures represented in the test, the related linguistic structures in DGS, and sign language acquisition studies were all presented. An important step within the adaptation process is to check the comparability of structures across sign languages. Consequently, for each separate linguistic structure, both the BSL structures and the DGS structures will be presented together in the following section on item representation. The original BSL items and their DGS counterparts will be presented in Tables 3.4 to 3.7.

#### 3.1.6.1 Spatial Verb Morphology

In the next step, all BSL items with spatial verb morphology were adapted to DGS. The adapted items are presented in Table 3.4, together with an indication of which aspects of spatial verb morphology they cover.

Table 3.4: Summary of the Adapted DGS Items for Spatial Verb Morphology

| *Item#* | *BSL item* | *Adapted DGS item* | *Type* |
|---|---|---|---|
| 2 | CAR ROW-ROW-ROW* | AUTO REIHE-REIHE-REIHE | Spatial verb (locative) with whole entity classifier |
| 5 | BOOK-ON | BUCH-AUF (BETT) | Spatial verb (locative) with whole entity classifier |
| 9 | BALL TABLE-ON | BALL TISCH-AUF | Spatial verb (locative) with whole entity classifier |
| 10 | TWO-PEOPLE-MEET | ZWEI-PERSONEN-TREFFEN | Spatial verb (motion) with whole entity classifier |
| 11 | DOG-IN | HUND-IN (KISTE) | Spatial verb (locative) with whole entity classifier |

| *Item#* | *BSL item* | *Adapted DGS item* | *Type* |
|---|---|---|---|
| 12 | PERSON-GO-DOWN-ESCALATOR* | PERSON-ROLLTREP-PE-RUNTER-FAHREN | Spatial verb (motion) with whole entity classifier |
| 13 | CHILD LOOK-UP | KIND SCHAUT-HOCH | Agreement verb |
| 15 | CAR-BEHIND | AUTO-HINTER (HAUS) | Spatial verb (locative) with whole entity classifier |
| 17 | BOX UNDER-BED | KISTE UNTER-BETT | Spatial verb (locative) with whole entity classifier |
| 18 | BOOK GIVE-TO-CHILD | (MUTTER) BUCH-KIND-GEBEN | Agreement verb |
| 20 | BOY-HIT GIRL-GET-HIT | JUNGE-SCHLAGENMÄD-CHEN-WIRD-GESCHLA-GEN | AB verb construction: body classifier, verb agreement, role shift |
| 27 | (BOY-left) POUR-WATER-OUT (BOY-right) WATER-POUR HAIR-WET | (JUNGE-links) WASSER-KOPF-GIESSEN AUF (JUN-GE-rechts) WASSER-KOPF-GIESSEN HAAR-NASS | AB verb construction: hand-ling classifier, verb agree-ment, role shift |
| 29 | MOTHER GIVE-LET-TER | MUTTER BRIEF-GEBEN | Agreement verb |
| 32 | MOTHER BOOK SHOWN-DOWN-WARDS | MUTTER BUCH-NACH-UNTEN-ZEIGEN | Agreement verb |
| 34 | DOG-IN-FRONT | HUND-VOR (KISTE) | Spatial verb (locative) with whole entity classifier |
| 38 | ROW-OF-CARS BOT-TOM-LEFT | REIHE-AUTOS-UN-TEN-LINKS | Spatial verb (locative) with whole entity classifier |
| 39 | DOG-LIE-INSIDE LEFT | HUND-LIE-GEN-INNEN-LINKS (KISTE) | Spatial verb (locative) with whole entity classifier |
| 40 | HOUSE-TOP-RIGHT | HAUS-OBEN-RECHTS | Spatial verb (locative) with whole entity classifier |
| 44** | teddy under bed | TEDDY-UNTER BETT | Spatial verb (locative) with whole entity classifier |
| 45** | dog behind box | HUND-HINTER KISTE | Spatial verb (locative) with whole entity classifier |
| 48** | child in front of car | KIND STEHEN-VOR-AUTO | Spatial verb (locative) with whole entity classifier |

* These items also belong to the category of number and distribution
** Items 41–50 are new developed items for the adapted DGS test

#### 3.1.6.2 Size and Shape Specifiers

Using the BSL items as a basis, the items were adapted to DGS. Table 3.5 summarizes the SASS items of the BSL test and the adapted DGS items.

Table 3.5: Summary of the Adapted DGS Items for Size and Shape Specifiers (SASS)

| *Item#* | *BSL item* | *Adapted DGS item* | *Type* |
|---|---|---|---|
| P3 | TEDDY-SMALL | TEDDY-KLEIN | indicating size |
| 16 | CURLY-HAIR | HAAR-LOCKIG | handshape, indicating "curliness" of hair by the movement |
| 21 | PENCIL THICK | BLEISTIFT DICK | handshape and non-manually |
| 22 | THICK-STRIPES-DOWN-TROUSERS | BREITE-STREIFEN-NACH-UNTEN-HOSE | handshape, movement, and non-manually |
| 46* | sweater with rows of dots | REIHEN-MIT-PUNKTEN-PULLI | handshape, movement, and non-manually |
| 49* | small pencil | BLEISTIFT-KLEIN | handshape and non-manually |

* Items 41–50 are new developed items for the adapted DGS test used in the main study

#### 3.1.6.3 Handling Classifiers

Handling classifiers represent how an object is held and manipulated. Therefore, these items are identical in BSL and DGS. This concerns Items 25, 35, and 37.

#### 3.1.6.4 Number and Distribution

These findings suggest that similar means are used in British and German Sign Languages to express number and distributive aspect. The results of the review of the DGS research on number and distribution are summarized in Table 3.6.

Table 3.6: Summary of the Adapted DGS Number & Distribution Items

| *Item#* | *BSL item* | *Adapted DGS item* | *Type* |
|---|---|---|---|
| 1 | LOTS APPLE | APFEL VIELE[25] | C-noun & quantifier |
| 2 | CAR ROW-ROW-ROW* | AUTO REIHE-REIHE-REIHE | (Noun)** & reduplication of classifier handshape |
| 6 | ONE-TEDDY | EIN-TEDDY | (Noun) & classifier handshape (singular) |
| 12 | PERSON-GO-DOWN-ESCALATOR* | PERSON ROLL-TREPPE-RUNTER-FAHREN | (Noun) & classifier handshape (singular), spatial orientation and path |
| 13 | FEW-CUPS | EINIGE-TASSEN | (Noun) & reduplication of classifier handshape (in space) |
| 24 | QUEUE | SCHLANGE-LEUTE | (Noun) & classifier handshape ("Up-right-4", with plural meaning) |
| 41*** | many cars | AUTO VIELE | C-noun & quantifier |
| 42*** | few pencils | EINIGE-STIFTE | (Noun) & reduplication of classifier handshape (in space) |
| 47*** | one ball | EIN-BALL | M-noun (singular) |

* These items also belong to the category of spatial verb morphology.

** Nouns are not signed in the test stimuli.

*** Items 41–50 are newly developed items for the adapted DGS test.

### 3.1.6.5 Negation

Based on the review of the literature on DGS as presented in the previous chapter, specific meanings of DGS negation for the items were identified. This was done in collaboration with a Deaf adult who also signed the test materials. The results are presented in Table 3.7. These results were the

[25] While filming the test items, the Deaf collaborator who had modeled the test materials signed VIELE in a post-nominal position. This was not noticed by the researcher reviewing the filmed test items. Also in Pilot 2 with five Deaf adults, they did not comment on the post-nominal position of the quantifier VIELE. This suggests that maybe the position of the quantifier seems to be flexible to a certain degree (maybe depending on the context), with a strong preference for a pre-nominal position. Yet, a post-nominal position is also allowed. That the Deaf adults did not notice or comment on the post-nominal position of the quantifier might also be influenced by the fact that the test items constitute decontextualized language use. But it could also be that less is known about the position of the quantifier VIELE in DGS than was previously assumed.

basis for the adapted DGS version. One specific sign, NICHT1, is articulated with a side-to-side headshake and was used throughout the test (see also Papaspyrou et al., 2008).

Table 3.7: Summary of Adapted DGS Negation Item

| *Item#* | *BSL item* | *Adapted DGS item* | *Negation sign** | *Verb (modification)* | *Non-manual marker* |
|---|---|---|---|---|---|
| 3 | ICE-CREAM NOTHING | EIS NICHTS1 | NICHTS1 | | Headshake |
| 4 | NOT-LIKE EAT | MAG-NICHT ESSEN | | Modification of modal MÖGEN | Headshake, facial expression |
| 8 | HAT NOTHING | HUT NICHTS1 | NICHTS1 | | Headshake |
| 23 | NOT-SLEEPING | NICHT-SCHLAFEN | | SCHLAFEN | Headshake (after the verb) |
| 28 | HEADPHONE NOTHING | KOPFHÖRER NICHTS1 | NICHTS1 | | Headshake |
| 30 | CHILD COAT RAIN NOTHING | KIND MANTEL KEIN REGEN | | REGNEN | Headshake, facial expression |
| 31 | CAN'T REACH | NICHT-HERAN-KOMMEN | | HERAN-KOMMEN | Facial expression, headshake (after the verb) |
| 33 | DOG NO COLLAR EAT BIG BONE | HUND NICHTS1 HALSBAND ESSEN KNOCHEN-GROSS | NICHTS1 | | Headshake, facial expression |
| 35 | NOT-DROP-CUP | NICHT-FALLEN-LASSEN GLAS | | Glas-NICHT-FALLEN-LASSEN | Headshake, facial expression |
| 36 | BALL NOTHING | BALL NICHTS1 | NICHTS1 | | Headshake, facial expression |

* Gloss NICHTS1 derived from Heßmann 2001a

#### 3.1.6.6 Noun/Verb Distinction

The four items in this category were adapted in the first version of the DGS test, but were then removed from the item pool following Pilot 2 (see below for results on Pilot 2). The decision to remove these items following Pilot 2 and not before was that the research literature on DGS is not comprehensive regarding this structure. Therefore, it was decided to also obtain feedback from the Deaf adults in Pilot 2 about this morphological noun-verb distinction in DGS. Based on that feedback and on existing research indicating that noun-verb pairs are not derivationally related in DGS (Becker, 2003), items testing the noun-verb distinction were removed.

### 3.1.7 The Test Materials

The test materials were (1) signed videos, and (2) pictures in the form of drawings that were appealing to the children and easy to recognize. The original test stimuli of the BSL test were used for this study. Some pictures were changed either for cultural reasons or to strengthen the focus on what the pictures had originally intended to depict. The pictures were simple, colored drawings. They are very child-friendly and were designed to focus on the essentials while avoiding any potential information-overload that could distract children from the intended task. The pictures depict easily recognizable objects and relationships, and are appealing to children in the targeted age range (3–11 years). Additional distractor items were included to reduce guessing, and the location of the target picture on the page was randomized.

As for lexical variation, it was decided together with the Deaf collaborator who had modeled the test materials to use one single variant throughout the entire test since the collected data of Pilot 1 did not reveal clear regional variation, and because the creation of three to four different test versions would have been complicated. The variant was decided by the Deaf person. If a child did not know these signs/variants in the vocabulary check, she/he was familiarized with them in a brief training session that followed the vocabulary check before the actual receptive skills test. In this training session, the test administrator also asked the child if she/he knew and understood the sign.

### 3.1.8 Test Instructions

The test administrator did not perform the test instructions live. Instead, the test instructions were in a standardized format as part of the computer-based test. The rationale for not performing instructions live was to ensure as highly standardized a testing situation as possible for all participants. The test starts with general instructions on video in DGS, followed by the vocabulary check. Instructions for other parts of the test follow later, also in DGS.

### 3.1.9 Test Software

An experienced programmer was contracted to develop a test interface that would fulfill the following criteria:

(1) Include as many features as necessary, but without overloading the interface

(2) Provide a user friendly interface for children, that is, it should be very easy in terms of navigation

(3) Provide flexibility for the test administrator, so that the she/he has the option of changing the order and the number of the items, and the number of times a test item or a set of instructions can be watched, etc.

The first version of the software was completed in September 2005. The test consisted of three sections. Before the first section, started the test administrator was asked to enter an ID for the children so that the results could be saved in a labeled file.

The three sections were (1) general introduction and test instructions, followed by the vocabulary check (Figure 3.4) where the test administrator had to mark the vocabulary on a checklist that had been adapted from the BSL test (Appendix E-1); (2) a training session where the Deaf signer in the video teaches the children the four lexical signs used in the DGS Receptive Skills Test (Figure 3.5), for which regional variants were identified during Pilot 1; (3) The Receptive Skills Test, which had an introduction followed by three practice items. In this first test version, the video was always shown on the left side of the computer screen with four buttons for navigation (*rewind*, *play*, *stop*, *forward*). The video did not start automatically.

Figure 3.4: Examples of Vocabulary Check of Adapted DGS Receptive Skills Test

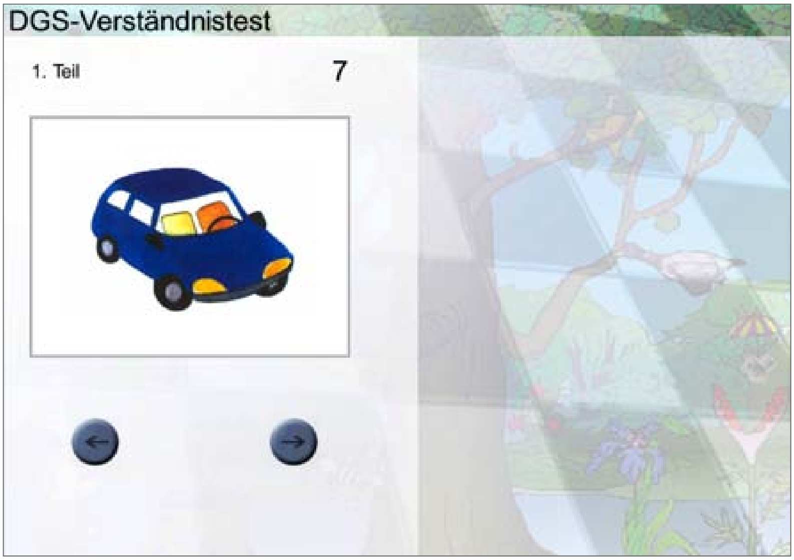



Figure 3.5: Example of the Training Session of the Adapted DGS Receptive Skills Test

In this version of the test, the child had to use the rewind button before she/he could re-watch the test item. After the child watched the test item, the three or four pictures (answers) faded in on the right side of the screen. The child could then choose the picture that represented the best (correct) answer. The child could click more than once (in case she/he was undecided); only the last picture that was clicked on was saved as a test response. Once the child clicked on a picture, a button depicting an arrow indicating *next* faded in as a signal to proceed to the next item. This design was chosen in order to make sure, that an answer had been chosen before proceeding. Once the child had proceeded to the next item, it was not possible to return to the previous one. This first test version consisted of three practice items, followed by 50 test items. At the end of the practice items, the test administrator had to save the results on the hard disk and at the end of the test items, the results of the test items were manually saved in the selected folder. Figure 3.6 shows an example of the test's layout as it appeared to the user.

Figure 3.6: Example of the DGS Receptive Skills Test's Computer Interface (first version, Pilot 2)

## 3.2 Pilot 2: Testing of First Test Version

In this section, the results of Pilot 2 with Deaf adults and non-signing hearing children will be presented.

### 3.2.1 Pilot 2 with Deaf Adults

The goal of Pilot 2 with Deaf adults was to check to which extent they would agree on the items of the first version of the adapted DGS test. The pilot testing with Deaf adults took place in October 2005. The Deaf adults were contacted directly by the researcher. A total of five Deaf adults were tested. Four informants were from Northern Germany, and one from South-West Germany. All informants were required to sign a consent form (Appendix F-1) and to fill out a short background questionnaire detailing their sign language use and their contact with other Deaf people (Appendix F-2).

The age range of the informants was from 23 to 56 years old (*M* = 39;6). Four informants had hearing parents and one informant had Deaf parents. All five had attended a kindergarten and a school for the Deaf. Four had a background in teaching DGS. They all reported that they used DGS in a variety of settings, such as with their family, and/or friends, at Deaf club, and at work.

One pilot test session took place in the South-West region, and the remaining four in the Northern region. All informants were tested individually. The Deaf adults were informed about the general goal of the pilot. The test was displayed on a laptop computer. Notes were taken during testing in order to record feedback from the Deaf adults. All of the Deaf adults completed the entire test. Since the test results are stored automatically, specific focus was placed on gathering feedback from the Deaf informants about their views concerning the pictures and the signed stimulus sentences. The Deaf adults were also asked to explain the reasons for their response choice. At the end, both the feedback (qualitative; Appendix F-3) and the test data of the Deaf informants (quantitative; Appendix F-4) were analyzed.

On most of the items (42/53, 3 practice and 50 test items), the Deaf adults were in agreement. The 11 items where the informants disagreed were examined more closely: (1) Items where at least three out of five Deaf adults agreed and where no feedback was provided remained unchanged (Items P3, 37); (2) Items where three (or four) out of five Deaf adults agreed, but also where feedback was provided were revised (Items 2, 15, 30, 32, 34, 38, 45, 46, 48); (3) Items where all Deaf informants agreed, but where feedback was provided, were also considered for revision (Items 11, 39, 47, 49). The results of this analysis revealed the need for some changes and these were made in the test revision.

The results of the Pilot 2 revealed the need for changes to some pictures and video stimulus sentences. The main revisions identified by the Deaf informants can be summarized as follows:

(1) Phonological or morphological errors:
   (a) Wrong movement; for example, a wrong movement indicating *rows of cars* (Item 2)
   (b) Use of wrong classifier handshape; for example, *a sitting dog* (not lying, Item 34) or *size of dots on a sweater* (Item 46)

(2) Syntax: wrong sign order used; for example, Item 34 (DOG BOX-IN) was wrong, it should be BOX DOG CL-DOG-IN

(3) Non-manuals: the headshake for negation should be spread over the entire phrase (Item 30), not only accompany the negation sign

(4) Changes to pictures:
   (a) Sometimes the target picture and the some of the distractors were too similar (e.g., Item 32)
   (b) Non-content considerations: the color or the style of a particular picture was found to be unsatisfactory (Items 42, 49)

The qualitative data also revealed that the derivationally-related noun-verb items are not morphologically related in DGS, and were therefore removed from the item pool (also based on the review of literature, Becker, 2003). The complete overview of the results of the Pilot 2 and the changes made to the materials can be found in the appendix (Appendix F-3).

### 3.2.2 Pilot 2 with Non-Signing Hearing Children

In October 2005, a Pilot 2 was conducted with non-signing hearing children. The rationale for including non-signing hearing children in a pilot study was that since the target group of Deaf children was very small, it was preferable not to test them until the final version was ready. The objective for this pilot was to obtain information on the user-friendliness of the test interface: Is the test easy to navigate? Or are there any general problems in the structure of the test etc.?

Pilot 2 with the non-signing hearing children was conducted in a kindergarten in South-West Germany. Before the testing, the researcher sent questionnaires to the kindergarten to obtain information on such variables as age of child, contact with Deaf people, and knowledge of sign language, together with a parental consent form (Appendix F-5). The head of the kindergarten forwarded both the questionnaire and consent form to the parents. All returned questionnaires and signed consent forms were collected by the teacher and handed to the researcher on the day of the pilot.

The test was set up on a laptop computer in the teachers lounge. The children were brought in individually from their classroom. One of the teachers was present during the testing session for the reassurance of the children.

The first child completed the entire test. After the first child completed the test, the procedure was changed to reduce the time required because the task was too difficult. Thereafter children did Items 1–25 and 26–50 alternately.

All illustrations were also available as hard copies. The children were given the hard copy of an item and asked to describe what they saw in order to make the goal of the task clear before showing them the videos. During the test, they had the option to view items twice. After each item, the children were asked to point either on the laptop screen or on the hard copy to the picture matching the video.

Two additional children (in Zurich) completed all items of the test. A total of 13 children were tested. Three completed the entire test, the remaining ten only half of the test. The age range was from 4;8 to 7;10 (*M* = 5;8).

The observations made during the testing provided insights into the user-friendliness of the test interface and were the basis for revisions made to meet the needs of the target group of children (see below Chapter 3, Section 3.2.3).

### 3.2.3 Revision of First Test Version

Based on the results of Pilot 2 with Deaf adults, the following revisions were made[26]:

(1) Re-filming of 10 items (for a complete list of these items and why they were re-filmed, see Appendix F-3)
(2) Revisions of 9 pictures (Appendix F-3)

Observations made during Pilot 2 testing of the non-signing hearing children resulted in the following changes to enhance the user friendliness of the test for the target group:

(1) Simplifying video navigation: larger buttons, and changing the interface so that only a repeat click on *play* is needed to re-watch a video; the *forward* button was deleted (since with the replay change, it no longer had any function, see Figure 3.7)
(2) Enlarging the entire test interface so that it fills the 15” laptop screen
(3) Blocking the possibility of clicking on a picture while the video was running, ensuring that it was only possible to click after watching the video,
(4) Automatically saving the results of the practice and test items at the end of the test (separately)
(5) Including the option to save information on how often a video clip had been watched and which pictures had been clicked on before the final choice (settings for the test administrator)
(6) Offering the option to select for each child and for each part of the test how often a video could be watched (the child had the option of watching the practice item up to three times, and the test items twice).

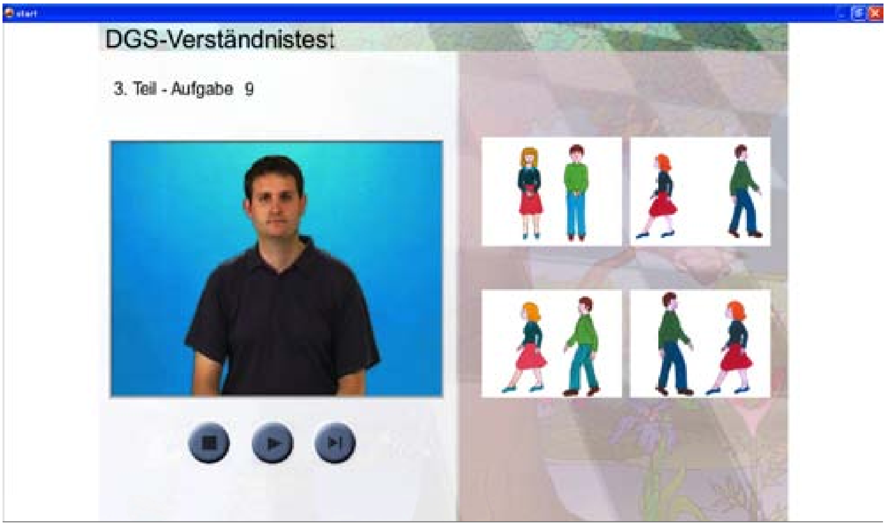
As stated earlier in Chapter 3, Section 3.1.6, due to input from Deaf informants, the four items that test morphologically-related noun-verb distinctions (three from the original test version and one from the newly developed items) were excluded for the subsequent testing (three practice items and 46 test items remained in the item pool). The numbering of items changed, following the deletion of the four items (Appendix F-6).

In the next section, the test procedures applied in the main study will be presented.

---

[26] Selected examples of the DGS Receptive Skills Test can be accessed in the Internet at http://www.signlang-assessment.info/index.php/german-sign-language-receptive-skills-test.html.

Figure 3.7: Example of the Revised Version of the DGS Receptive Skills Test (for Main Study)




## 3.3 Description of the Test Data from the Main Study

Different sets of data were collected during the main study: (1) background information from questionnaires; (2) the test results (raw scores); (3) a test protocol/observation sheet; and (4) video recordings made during the test session.

### 3.3.1 The Subjects

A total of 74 Deaf children from 3;9 to 10;10 years old (*M* = 7;0) were tested between February and June 2006. Of these 74, the raw scores of 20 children were excluded from the data analysis, because: (1) 14 did not complete the test or the test administrator stopped the test session after ten consecutive fails; and (2) 6 children were excluded because of a reported additional disability. Therefore, the final number of Deaf children in this study is 54 (29 male, 25 female).

The Deaf children came from 5 schools for the Deaf where some form of signing was used[27]. The form and degree of signing varied across the 5 schools: (1) implementation of a bilingual philosophy using DGS as language of instruction (1/5); (2) a bilingual pilot classroom with subsequent use of DGS in other classes across the school (2/5); and (3) signing used to a certain degree as means of instruction (2/5), ranging from DGS to manual communication, such as LBG (Signed German). The number of schools involved had to be limited according to the available resources. The schools were located in five different geographical regions in Germany (Northern, Eastern, Western, South-West, and Southern). Because of the limited number of potential subjects, the schools were asked to identify Deaf children of Deaf families with a native signing background. This procedure was chosen in order to obtain as homogeneous a group as possible for the main phase of the study. While some schools followed this instruction, others suggested testing all children in the age range from four to eight years old. It was decided to include all children in this study. The effects of this broader selection of participants will be considered for the data analysis.

Table 3.8 shows the Deaf children across schools and regions, including parents' hearing status.

Table 3.8: Deaf Children Across Regions, Schools, and Parental Hearing Status ($N$ = 54)

| *Region* | *School* | *N* | *Deaf parent(s) (n)* | *Hearing parents (n)* |
|---|---|---|---|---|
| Northern | School #1 | 16 | 13 | 3 |
| Eastern | School #2 | 17 | 8 | 9 |
| Western | School #3 | 12 | 5 | 7 |
| South-West | School #4 | 3 | 3 | N/A |
| Southern | School #5 | 6 | 5 | 1 |
| Total | | 54 | 34 | 20 |

[27] In almost all institutions, the age range from 4-8 years old was covered (kindergarten and elementary school were housed in the same institution). In one region, the kindergarten and the elementary school were two different institutions, but in the same neighborhood. Since no comparison between institutions was made and because these two institutions followed a similar language philosophy, they were taken together for presentation purposes.

Reported hearing losses for these 54 children were: (1) one child with a mild hearing loss (25–40 dB); (2) two with a moderate hearing loss (40–70 dB); (3) 29 with a severe hearing loss (70–100 dB); and (4) 18 with a profound loss (> 100 dB). No information was provided for 4 children.

## 3.3.2 Educational Background of the Subjects

In addition to testing children's DGS skills, three sets of questionnaires were distributed to collect demographic background information. One questionnaire was given to the parents or legal guardians to obtain information on language use at home and the preferred language(s) of the Deaf child (Appendix G-1) Two questionnaires were completed by the teachers, one for each individual child (Appendix G-2) and a second one requesting general information about the school (Appendix G-3).

### 3.3.2.1 Parent Questionnaire

The parent questionnaire was designed to collect background information about the child. It comprised nine items, including information about date of birth, onset of hearing loss, and when the child first started to sign; the remaining items concerned the use of languages in different situations – with different family members or in contact with people outside of school. The questionnaire was comprised of yes-no questions (e.g., "Does the child have contact with anyone outside of school who signs?"), multiple-choice items, where the parents had to check off one or more answers (e.g., "Which language(s) and means of communication are used most frequently at home?"), and questions requiring the respondents to fill in specific information (e.g., "When did the child first start to sign?").

### 3.3.2.2 Student Questionnaire

The student questionnaire consisted of seven items and was completed by the teachers. Most of the items were similar to the parent questionnaire, with particular regard to languages used and means of communication at home. The final item was a rating scale for the teachers to provide their informal evaluation of the receptive and productive DGS skills of each participating Deaf child on a scale from 1 to 6, with 1 representing the best and 6 the worst, following the German school grading system.

#### 3.3.2.3 Teacher Questionnaire

The third questionnaire was also completed by the teachers. This instrument consisted of eight items, divided in two parts: The first part dealt with background information about the teacher, such as professional qualifications, hearing status, grade level taught; and the second part dealt with communication issues, such as the use of different means of communication used by children and teachers in different settings, as well as a self-evaluation of the teacher's receptive and productive sign language skills.

The rationale for using two questionnaires with similar content was to maximize the validity of the background information obtained for each child, that is, to obtain information about the children from both the parents and the teachers. This made it possible to cross-check the information provided by each group of respondents.

Teachers at the schools coordinated the distribution of all questionnaires. The teacher questionnaires were placed in the teachers' individual mailboxes at the schools, the parent questionnaires were sent home with the children. The parent questionnaires were accompanied by a cover letter explaining the purpose of this study and a consent form to be signed by the parent or legal guardian. In signing the consent form, the parent or legal guardian gave permission for the child to be tested, and for the child to be video-taped during the testing session (all videos were deleted after the analysis). Only children for whom a signed consent form was available were included in this study. The signed forms remained in the schools.

In order to ensure the privacy of the participating Deaf subjects, no names were used at any time during this study. To enable the parent and student questionnaires (and later the test scores) to be linked, the questionnaires received linked codes. As an example, in the code *E – C – 01*, *E* means that it is a parent questionnaire, *C* identifies one of the five test sites, and *01* identifies a specific subject. The parent questionnaires were distributed (via the children) by the teachers who had received training by this researcher on how to read the identification coding system. The parents returned the questionnaires to the teachers who then gave them to the researcher.

The identification coding system for the student questionnaires, completed by the teachers for each subject, was very similar though with a different first letter to identify it as a student questionnaire; the codes were matched with the parent questionnaires. Following state education departmental requirements, the information codes for the questionnaires and the testing remained with one designated teacher in each school, in order to ensure the protection of the personal data of the Deaf subjects. All questionnaires were developed in close collaboration with the participating schools. Feedback from the teachers led to several revisions of the questionnaires.

The procedure employed in this study to measure the variables specific to the research question will now be described.

## 3.4 Protocol of the Main Study

In the next section, the entire procedure from contacting the schools to the actual testing will be described.

### 3.4.1 Contacting the Schools and the Participants

The schools were contacted by mail. The researcher also traveled to four out of five school sites to introduce the proposed study in person. Once the school administration and the teachers agreed that they wanted to take part in this study, an application with a statement of purpose, questionnaires, consent form, etc. was sent to the states' education departments. Once approval of the study and materials had been given by the department of education, the process of recruiting participants began.

Participating children received a test information package from the teachers to give to their parents. The package contained a cover letter explaining the scope of the study, a consent form, and the parent questionnaire. In most cases, the schools added a letter stating support for the study. The parents or legal guardian were required to sign a form consenting to their child participating in this study, and then to return the signed consent form to the teachers/schools. Also, the teachers explained the purpose of the study to children who might be interested in taking part.

### 3.4.2 Time of Test Administration

The testing took place between February and June 2006. All testing sessions were conducted by the researcher.

A meeting was held with the teachers on the morning of the day of the test to set up a timetable for the testing of each child. Since the testing was done on a one-to-one basis, each participating child was taken individually from their classroom for the testing. They were collected and returned to the classroom by the researcher. The children participated voluntarily in this study. The children were also told that they could withdraw from the test at any time.

The session for the younger children (3;9–5;6) took approximately 30 minutes, and for the older children (> 5;7) approximately 20 minutes. The majority of test sessions took place in the morning, with only a few in the early afternoon (latest at 2.00 pm).

### 3.4.3 Test Location

At all the sites, the testing took place in a separate room located in a quiet part of the school building. In most cases, the testing room was a classroom or staff meeting room that was not in use at the time. In one school, testing took place in the computer lab. Test set-up and completion was facilitated by the fact that the test was administered on the researcher's large-screen laptop computer, thereby allowing some flexibility.

The test location was prepared by the researcher, with the laptop and all other testing material on one table. The computer and table were arranged to ensure there were optimal light conditions. Two chairs, one for the Deaf child and one for the researcher (in one school, specially-sized chairs for the younger children were available) were placed at the table and positioned in an angle, which allowed both the child and the researcher to be video-recorded with a single camera.

The age range of the Deaf participants from 4–8 years old straddles kindergarten and school. In four of the five regions, the kindergarten and elementary school were located in the same building. In one region, the kindergarten and the elementary school were two separate institutions in separate buildings in the same neighborhood (see also above, Chapter 3, Section 3.3.1).

No major technical problems were encountered during the testing. All test scores were saved on the laptop hard-drive as well as backed up on an external hard drive. For the video-recordings, mini DV-tapes were used.

### 3.4.4 Testing Protocol

When the children entered the testing room, they were shown where the testing would take place. The younger children required a few minutes to familiarize themselves with the setting and the room. Before the test session started, the researcher started to communicate with the child in DGS about something unrelated to the test as a warm-up (e.g., asking them if they were familiar with the use of a laptop). The use of the mouse was explained to those children unfamiliar with computers. Two types of computer mice were available from which children could pick one: (1) a wireless-mouse for adults; and (2) some specially designed mice for children. After a few trials, the mice made for children were removed since they did not work properly, and only the wireless-mouse was then used. Some children required a few minutes to familiarize themselves with the use of the mouse. Fourteen out of 54 children were not able to use the mouse at all. In these cases, the children were allowed to select the answer by pointing at printed copies of the pictures or at the computer-screen. Whenever this happened, the experimenter used the mouse to enter the answers based on those selected by the children. In such cases, the arrow of the mouse was always kept pointed to the side of the computer-screen where the pictures were not displayed in order to avoid the child being influenced by the position of the mouse.

During the testing session, the researcher sat beside the Deaf child, but a little bit back so as to avoid communication (except at the beginning of the session) but to still be available to answer questions. All test instructions were provided in video format. Some younger children with limited language ability found it difficult to understand the test format despite the instructions. In these cases, the children were encouraged to go ahead and start with the practice items. This approach worked well, since it is quite easy to navigate through the test (click on the play button of the video first, select a response, and then go to the continue button).

At the end of each test session, the results were saved to an individual folder on the computer's hard drive.

An observation sheet (Appendix G-4) and a scoring sheet for the vocabulary check were also used. The observation sheet included the date and time of the testing, and the child's gender and ID. It was also possible to indicate whether the child used or did not use the mouse by himself/herself, and if she/he pointed at the screen, the printouts, or both. In addition, space for additional observations was provided.

The scoring sheet for the vocabulary test (Appendix E-1) included information about the test date and time and child ID, and whether the child knew the sign or not. Since the complete session was videotaped, the score sheet of the vocabulary check could be double-checked with the recording. When a child did not produce a sign corresponding to the vocabulary item, the experimenter asked them what they saw. This way it could be ensured that the children knew the vocabulary. When this approach did not work, it was possible to skip one item and return to it later in order to be sure that all children knew the 21 vocabulary items of the vocabulary check.

The next section will introduce the analysis of the data; the complete analysis will be presented in the next chapter.

## 3.5 Data Analysis

The present study examines methodological and theoretical issues involved in the adaptation from the BSL Receptive Skills Test (Herman et al., 1999) into DGS, with an emphasis on linguistic and cultural aspects and on psychometric properties. Only those research questions that require statistical or other analyses will be presented here. Theoretical questions will be discussed in Chapter 5 ("Discussion").

The study is guided by the following research questions:

1 Does the adapted DGS test provide evidence of having sound psychometric properties?
   - 1.1 Item analysis: Does the adapted DGS test show evidence of item facility and discrimination index?
   - 1.2 Fit of newly developed items: How do the newly developed items fit into the adapted test?
   - 1.3 Distractor analysis: Does the distractor analysis show evidence of the effectiveness of the distractors?
   - 1.4 Does the test show evidence of homogeneity?
   - 1.5 Does the test show evidence of reliability (e.g., Cronbach's alpha)?
   - 1.6 Does the test offer evidence of relations to an external variable (e.g., teachers' ratings of the children's sign language skills)?

1.7 Does the test show evidence of content validity?[28]

2 What are the relationships between the Deaf children's raw scores and other variables (gender, age of sign language exposure, parental hearing status, chronological age)?
   2.1 Does the gender of the children have an impact on their test performance?
   2.2 Does the age of sign language exposure have an impact on children's test performance?
   2.3 Does parental hearing status have an impact on children's test performance?
   2.4 Does chronological age (in the subgroups of Deaf children of Deaf parents, and Deaf children of hearing parents) have an impact on children's test performance?

In the following section, the analyses used to address each research question (i.e., using statistical analysis of the test results) will be briefly described.

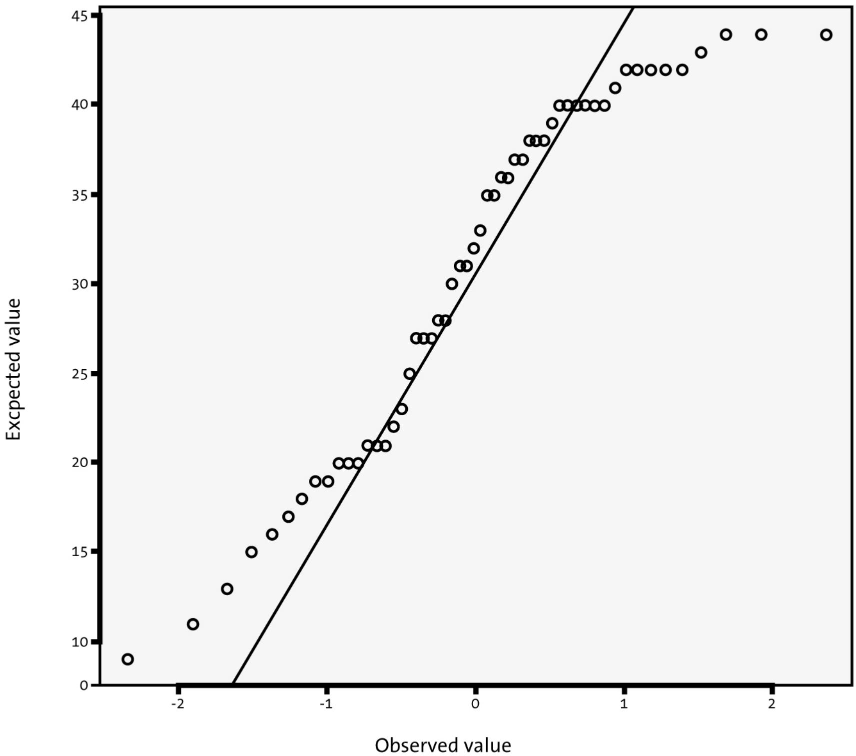
### 3.5.1 Statistical Assumptions

Two tests were used in order to determine the normal distribution of the sample. For normally distributed samples, parametric statistical testing methods can be applied. For non-normally distributed samples, nonparametric statistical testing procedures should be applied (Kiess, 1996). Using a histogram of the raw score variable with a normal curve overlaid (Appendix H-1) it can be seen that the sample is left skewed and thus does not represent a normally distributed sample ($M = 30.72$, $SD = 10.15$, $N = 54$). Additional support for the non-normal distribution was found using a Q–Q Plot (Figure 3.8), in which a diagonal line represents normally distributed scores and dots the observed scores in the study. In a normally distributed sample, the observed scores would be closer to the diagonal line (for descriptive statistics on the raw score variable see Appendix H-2).

---

[28] This research question is more a theoretical/review-based question than an empirically-based question in this study and will be investigated in Chapter 5 "Discussion".

Figure 3.8: Normal Q–Q Plot of the Variable Raw Score (*N* = 54)




The same observation was confirmed on the histogram and the normal Q–Q Plot for the variable Age (Appendix H-3, H-4). Based on these results, it was decided to use nonparametric testing procedures for all statistical procedures related to research questions (2.1) to (2.4).

It was decided to use an alpha level of .05 (2-tailed) as the level of statistical significance because of the small sample size and the rather new area of investigation.

In addition, it was decided to follow Bortz's (1999) and Cohen's (1992) proposals for determining the effect size of a correlation coefficient of (1) .10 as small, (2) .30 as medium, and (3) .50 as large.

Since no data of the BSL Receptive Skills Tests were available, it was not possible to run any (direct) statistical comparisons between the BSL and the adapted DGS test. However, some issues regarding theses two tests will be discussed in Chapter 5 (“Discussion”).

The statistical package SPSS (Statistical Package for the Social Sciences) was be used for the analysis (e.g., Gaur & Gaur, 2006).

### 3.5.2 Item Analysis

As a first step in conducting an item analysis of the adapted DSG test, the item facility $p_i$ and discrimination index $r_{it}$ will be calculated (Rust & Golombok, 2000).

#### 3.5.2.1 Item Facility

Item facility refers to the degree to which the respondents, taken as a group, get a particular item right or wrong (Osterlind, 2001). Item facility can be calculated by a simple formula of number of subjects divided by the number of correct answers for each item.

#### 3.5.2.2 Item Discrimination

Item discrimination distinguishes items as being at different levels of abilities. For example, difficult items should be solved by subjects who achieved a high score on a test; however, if subjects with a low average score solve purportedly difficult items, then there may be a problem with those items.

### 3.5.3 Newly Developed Items

Based on the item analysis, the item facility and discrimination index for the 10 newly developed items were calculated in order to see how they fit into the overall test.

### 3.5.4 Distractor Analysis

To examine the quality of the distractors used in this study, a facility index and a discrimination index was calculated for each distractor of each item, applying the same statistical procedures as for the item analysis.

### 3.5.5 Homogeneity of the Test

Investigating the homogeneity of a test is a common procedure in test development (Fisseni, 2004). The goal of a test is to measure a certain trait. The different items in a test should measure different facets of the same trait, with the result that there should be an overlap between these facets. This overlap between facets of a test is the homogeneity. In order to investigate the homogeneity, an inter-item correlation was applied.

### 3.5.6 Evidence for Reliability

Reliability refers to whether a test actually measures what it is intended to measure (Rust & Golombok, 2000). The internal consistency of a test is measured by statistical analyses, such as a split-half analysis or reliability coefficient (Rust & Golombok, 2000). In order to investigate the internal consistency of the measure of this study, the reliability coefficient of Cronbach's alpha was calculated.

### 3.5.7 Evidence Based on Relationships with Other External Variables

In order to investigate whether there is evidence for a relationship to other variables that can provide support for the validity of the adapted DGS test, the possible relation of the children's test performance was compared with an external variable. This external variable is the teachers' ratings of the Deaf children's receptive and productive DGS skills. For these ratings, provided in the teacher questionnaires, scales were used ranging from 1 to 6 (1 being the highest performance). The nonparametric Spearman rank correlation coefficient $r_s$ between the children's raw scores and the teachers' rating was performed.

### 3.5.8 Evidence for Validity

Since the evidence of validity is review-based, the issue of content validity based on studies of DGS and acquisition studies will be presented in Chapter 5 ("Discussion").

Having outlined the procedures that were performed to examine the psychometric properties of the adapted DGS test, the type of data analyses that were performed to address questions of the relationship between the Deaf children's test performance and other variables, such as gender, age of sign language exposure, hearing status of the parents, and chronological age will be described.

### 3.5.9 Test Performance of Deaf Children

In order to investigate questions about the test performance of the Deaf children in relation to variables of (1) gender, (2) age of sign language exposure, (3) parents' hearing status, and (4) chronological age, a different set of statistical procedures were applied. Since the whole sample is not normally distributed, nonparametric testing procedures such as the Mann-Whitney U Test to compare groups, the Spearman rank correlation coefficient $r_s$ for correlations were applied, and a regression model will be applied.

In the next chapter, the results of the study will be presented, guided by the research questions that motivated this study.

# 4 Results

In this chapter, the results of the adaptation of the BSL Receptive Skills Test to DGS will be presented. This chapter is divided into three main sections: (1) description of the sample; (2) presentation of the results, addressing the different research questions; and (3) a summary of the results.

## 4.1 Description of the Sample

A total of 74 Deaf children participated in this study. Only the data of 54 children were used for the statistical analysis addressing the different research questions. The data of the remaining 20 children were excluded for the data analysis, because (1) 14 did not complete the test or the test administrator stopped the test session after ten consecutive fails, and (2) 6 children were reported to have an additional disability.

Thirty-four (63%) of the Deaf children came from Deaf families with at least one Deaf parent, and 20 Deaf children (37%) came from hearing families. The whole sample consisted of 29 male and 25 female Deaf children between 3;9 and 10;10 years (*M* = 7;0) who attend one of five schools or kindergarten programs that either (1) implemented a bilingual philosophy using DGS as the language of instruction, (2) a bilingual pilot classroom with subsequent use of DGS in other classes across the school, or (3) used signing to a certain degree as the means of instruction, ranging from DGS to manual communication, such as LBG. Table 4.1 provides a descriptive overview of the whole sample.

Table 4.1: Description of the Sample (*N* = 54)

| *Parents' hearing status* | *Male subjects (n)* | *Female subjects (n)* | *Age range (M)* |
|---|---|---|---|
| Deaf parents (*n* = 34) | 19 | 15 | 3;9–10;10 (6;10) |
| Hearing parents (*n* = 20) | 10 | 10 | 5;2–9;6 (7;4) |
| Total | 29 | 25 | 3;9–10;10 (7;0) |

For the research questions regarding quality of the test instrument (research questions 1.1. to 1.6.), only the subgroup of Deaf children of Deaf parents were included in the data analysis. Although the test targets all Deaf children (those with hearing parents as well as those with Deaf parents), it is important to have as homogeneous a sample as possible with early access to DGS in order to adapt and further develop a test that really does tap DGS development. The performance of a group of native signers provides a model against which the performance of children with other types of language exposure can be measured. A similar approach was used during the development of the BSL Receptive Skills Test (Herman, Holmes, & Woll, 1998). The second set of research questions (2.1 to 2.4), addressing the test performance of the Deaf children, was analyzed with data from the whole sample or by comparison of both subgroups.

The socio-demographic information used in this present study is based on the questionnaires filled out by the teachers, as introduced in Chapter 3. Three questionnaires were used in the study; two questionnaires were filled out by the teachers and one by the parents (see Chapter 3, Section 3.3.2). The decision not to use the questionnaire filled out by the parents was made because (1) it was obvious from the returned questionnaires that parents whose first language was not German had difficulties comprehending the questionnaire, and (2) for this same reason, in one of the five schools, the teachers did not hand out the questionnaires to the parents in the first place, but rather, filled them out themselves to the best of their knowledge. Therefore, only the questionnaires filled out by the teachers were used.

In Table 4.2, an overview of all languages and means of communication used in the children's home is presented.

Table 4.2: Languages Used in the Children's Home (Multiple Responses are Not Shown) ($N$ = 54)

| *DGS (n)* | *Other sign languages (n)* | *German (n)* | *LBG (n)* | *Home signs (n)* | *Two Slavic languages (n)* | *Three other European spoken languages (n)* | *Three other spoken languages (n)* |
|---|---|---|---|---|---|---|---|
| 34 | 5 | 31 | 19 | 11 | 3 | 8 | 3 |

In the following sections, data from the subgroup of Deaf children of Deaf parents and the subgroup of Deaf children of hearing parents will be presented.

### 4.1.1 Deaf Children of Deaf Parents

This subgroup consisted of 34 children (19 male, 15 female). The age range was from 3;9–10;10 years (*M* = 6;10). The educational background questionnaire completed by the teachers provided information about the home languages of the children. Of these 34 families, 11 used one sign language at home (8 families DGS, 2 families DGS and LBG, 1 family another sign language), whereas the remaining 23 families used at least two languages at home (mostly DGS in combination with German, with another sign language, or with another spoken language). Thirteen of these 23 families also used LBG (Signed German). Of these 23 families (1 family with 2 children), two families did not use a sign language but used two spoken languages and LBG. Five families also used home signs.

Table 4.3 summarizes the languages and means of communication used in the children's home of the subgroup of Deaf children of Deaf parents. The cells in the top row present the different categories in the use of different languages. This is supplemented in the left-hand column by categories that include (additional) information on the use of LBG, home signs, and German.

Table 4.3: Home Language Use of Deaf Children of Deaf Parents (*N* = 34)

| | *One sign language (n)* | *Two sign languages (n)* | *One sign & one or two spoken languages* (n)* | *One or two spoken languages and LBG (n)* | |
|---|---|---|---|---|---|
| | 11 | 1 | 4 | 3 | |
| LBG / home sign (*n*) | | | 11 | | |
| German and LBG / home sign (*n*) | | 4 | | | |
| Total | 11 | 5 | 15 | 3 | 34 |

*One child with two spoken languages

### 4.1.2 Deaf Children of Hearing Parents

The subgroup of Deaf children of hearing parents consisted of 20 children (10 males, 10 females). The age range was 5;2–9;6 years (*M* = 7;4). Five of these 20 families used only one language or means of communication (LBG) at home (1 family DGS, 1 spoken German, 2 another spoken language, 1 LBG). Of the remaining 15 families, 5 used at least one spoken or a sign language (German, another spoken language, or DGS) together with home signs and/or LBG at home. The remaining 10 families used at least two languages at home: 2 families used one sign language (DGS) and one spoken language (German); 8 families used 2 spoken languages (various combinations of German with another spoken language), sometimes in combination with LBG and/or home signs.

Table 4.4 summarizes the languages and means of communication used in the children's home of the subgroup of Deaf children of hearing parents. The cells in the top row present the different categories in the use of different languages. This is supplemented in the left-hand column by categories that include (additional) information on the use of LBG and home signs.

Table 4.4: Home Language Use of Deaf Children of Hearing Parents (*N* = 20)

| | *One sign language (n)* | *LBG (n)* | *One sign & one spoken language (n)* | *One spoken language (n)* | *Two spoken languages (n)* | |
|---|---|---|---|---|---|---|
| | 1 | 1 | 1 | 4 | 3 | |
| LBG / home sign (*n*) | 1 | | 1 | 3 | 5 | |
| Total | 2 | 1 | 2 | 7 | 8 | 20 |

Having presented the whole sample and the two subgroups, the investigation of the research questions will be presented in the following section.

### 4.1.3 Examining the Research Questions

Methodological and theoretical issues involved in the adaptation of the BSL Receptive Skills Test to DGS have been investigated with an emphasis on linguistic and cultural issues and the psychometric properties of the adapted DGS test. Only the results of the empirically driven research questions will be presented here. The data analyses addressing research questions (1.1) to (1.6) regarding the quality of the test instrument were performed on the subgroup of Deaf children of Deaf parents only. The data analyses addressing research questions (2.1) to (2.4) were performed on the whole sample or through comparison of the two subgroups. The study was motivated by the following research questions:

1. Does the adapted DGS test provide evidence of having sound psychometric properties?
   - 1.1 Item analysis: Does the adapted DGS test show evidence of item facility and discrimination index?
   - 1.2 Fit of newly developed items: How do the newly developed items fit into the adapted test?
   - 1.3 Distractor analysis: Does the distractor analysis show evidence of the effectiveness of the distractors?
   - 1.4 Does the test show evidence of homogeneity?
   - 1.5 Does the test show evidence of reliability (e.g., Cronbach's alpha)?
   - 1.6 Does the test offer evidence of relations to an external variable (e.g., teachers' ratings of the children's sign language skills)?
   - 1.7 Does the test show evidence of content validity?[29]
2. What are the relationships between the Deaf children's raw scores and other variables (gender, age of sign language exposure, parental hearing status, chronological age)?
   - 2.1 Does the gender of the children have an impact on their test performance?
   - 2.2 Does the age of sign language exposure have an impact on children's test performance?
   - 2.3 Does parental hearing status have an impact on children's test performance?
   - 2.4 Does chronological age (in the subgroups of Deaf children of Deaf parents, and Deaf children of hearing parents) have an impact on children's test performance?

[29] This research question is more a theoretical/review-based question than an empirically-based question in this study and will be investigated in Chapter 5 "Discussion".

The results are organized in the order of these research questions, with each question motivating a separate section of this chapter.

### 4.1.4 Item Analysis of the Adapted DGS Test

In the following section, the results of the item analysis, which consists of the item facility value $p_i$ and discrimination coefficient $r_{it}$, will be presented. The item analysis was performed only with the subgroup of Deaf children of Deaf parents. A total of 49 items of the original 53 (3 practice and 50 test items) of Pilot 2 (see Chapter 3, Section 3.2.3) will be analyzed here. Four test items were removed because they represent a linguistic structure (noun-verb distinction) that does not exist in DGS. The results of the item analysis contributed to whether items will be removed from the item pool for subsequent analysis (or suggested for revision with subsequent new piloting for a standardization study).

#### 4.1.4.1 Item Facility

The value of item facility can range from $p_i$ = -1.0 to +1.0. The greater the number of respondents who get a particular item right, the higher the investigated value (e.g., $p_i$ = .90) (i.e., the easier the item). When fewer participants get an item right, the value of item facility is lower (e.g., $p_i$ = .20) and therefore the item is more difficult. A large number of items scored as very easy or very difficult is not desirable since this tells us little about varying levels of language abilities within a given population (Alderson et al., 1995; Osterlind, 2001).

In the original BSL Receptive Skills Test (Herman et al., 1999), the items are ordered across different linguistic categories and according to their difficulty. The goal of the adapted DGS Receptive Skills Test is to test the development of specific structures of morphology and syntax in DGS. Therefore, the facility value was investigated (1) to find out the degree of difficulty (with the goal of ordering the items of the adapted DGS test in terms of difficulty in the revised version), and (2) to remove items which were too easy or too difficult since they do not differentiate between individuals and therefore do not contribute to the goal of the test. The statistical package SPSS was used to obtain the facility value (e.g., Gaur & Gaur, 2006).

The results of the item facility will be presented together with the item discrimination coefficient.

#### 4.1.4.2 Item Discrimination

The item discrimination coefficient differentiates between subjects with a generally high vs. a generally low test score. A high discrimination coefficient indicates that subjects with a high overall test score are likely to get a specific item right and subjects with a low overall test score are likely to get that item wrong (Lienert & Raatz, 1998; Osterlind, 2001; Rust & Golombok, 2000).

The discrimination coefficient $r_{it}$ can range from -1.0 to +1.0. The ideal value for a discrimination index is $r_{it} = 1$. A discrimination index of $r_{it} = 0$ on an item indicates that it has been answered by participants with both high and low overall scores. Such items should be deleted as they do not contribute to the overall test. A negative discrimination value indicates that participants with low overall scores tend to get that item right, whereas participants with a high score tend to get that item wrong. These items should also be revised or omitted (Fisseni, 2004).

The facility value together with the item discrimination index should help the test designer decide whether items need to be excluded or revised. The scatter of the facility index $p_i$ for individual items (e.g., from .25 to .90), indicates the degree of homogeneity among these items. A lower $p_i$ value indicates greater homogeneity, but in order to allow for the possibility of differentiation (in the case of this test differentiation by age), items with $p_i > .50$ and $< .50$ will be retained in the item pool (Fisseni, 2004). A biserial point correlation was applied to calculate the discrimination index for dichotomous coded variables (Alderson et al., 1995; Rust & Golombok, 2000). A Pearson product-moment correlation can be used, with the score for each item and the corrected score for the full test (whole-part correction) (Diehl & Staufenbiel, 2002).

#### 4.1.4.3 Results of the Item Analysis

Items were retained in the item pool when they met the following criteria (Fisseni, 2004; Lienert & Raatz, 1998): (1) items with facility value $p_i$ between .25 to .90; and (2) items with an item discrimination coefficient $r_{it} > .25$.

A total of 10 items were removed from the item pool for subsequent analysis (Appendix I-1) as they did not meet these criteria[30]. In relation to facility value, 1 item was removed because of a very high $p_i = .98$ (Item 38) and 4 items were removed because of a $p_i < .25$ (Items 31, 37, 41, 44).

Nine items with a discrimination index $r_{it} < .25$ were discarded (Items 2, 14, 31, 35, 36, 37, 41, 42, 44) as failing to meet the discrimination index criteria. Of these nine items, five had a value $p_i$ within the defined range of $p_i > .25$ to $< .90$ (Items 2, 14, 35, 36, 42); four items failed on both criteria, having a facility value $p_i$ not within in the defined range, and with a $r_{it} < .25$ (Items 31, 37, 41, 44). Some of the easier items (e.g., $p_i = .87$) were retained for use in the revised version as "icebreaker" items at the beginning of the test.

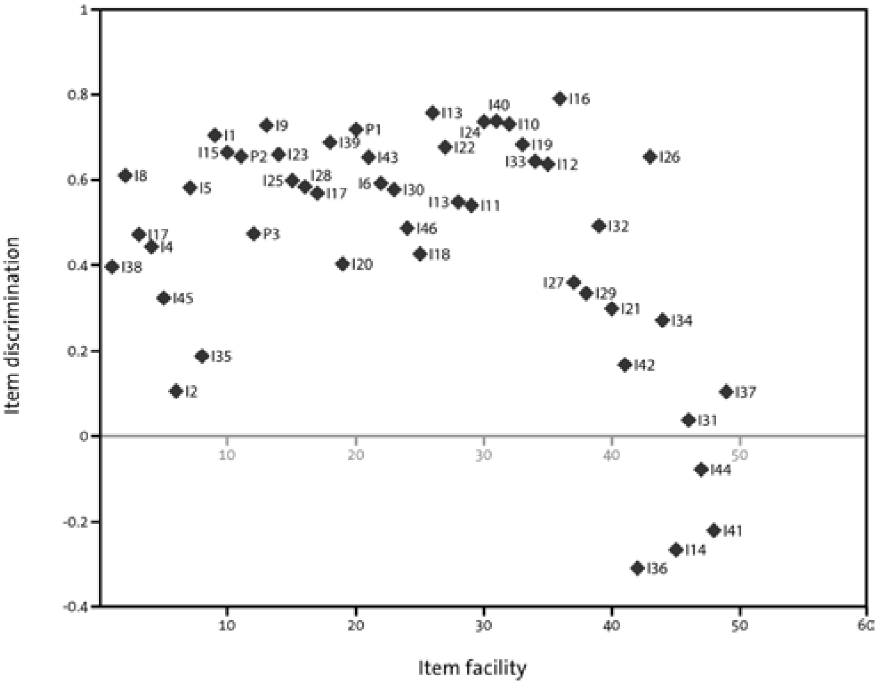
Items having a facility value within the defined range of $p_i = .25$ to .90 but with a $r_{it}$ below .25 were excluded. Items with negative values are those items on which participants with a low overall raw score perform well; these needed to be discarded or revised (Items 14, 36, 41, 44). In sum, a total of 10 out of 49 (3 practice items and 46 test items) were discarded from the item pool for subsequent analysis. Some items might be considered for revision, requiring new piloting (see Chapter 5).

Item difficulty and discrimination index are empirically related. When $p_i$ is low or high, $r_{it}$ will also be low or high (Fisseni, 2004). The interdependence of the facility value, discrimination coefficient, and homogeneity of the test and test items will be presented later (see Chapter 4, Section 4.1.6).

Figure 4.1 displays each item together with its facility value and discrimination index, and thus provides information on the relationship between facility and discrimination for each item. Many of the test items have a high facility value, and are thus easier.

[30] These items will be removed for subsequent analyses. But some of these items might be suggested for revision for the standardization study, which would include new piloting before standardization is possible. The reason for revision is content-driven since some of the items represent important linguistic structures which are relevant for inclusion in a DGS test that targets Deaf children's comprehension of morpho-syntactic structures ages 4-8 years old. This issue will be further investigated in Chapter 5 ("Discussion").

Figure 4.1: Interdependence of Facility Value and Discrimination Index of the Test Items




Having described the results above, now the fit of the newly developed items will be presented (items which were not adapted from the BSL test).

#### 4.1.4.4 Fit of the Newly Developed Items

Ten additional items were developed in the course of the adaptation. Data relating to nine of the ten will be presented here. One item had been removed from the item pool after the Pilot 2 study (see Chapter 3, Section 3.2.3) because it represented a linguistic structure (noun-verb distinction) that does not exist in DGS. Of these nine items, four were suggested for removal (or revision) based on item analysis (Items 38, 41, 42, 44). Item 38 was too easy ($p_i = .97$, $r_{it} = .396$); and Items 41, 42, and 44 had a discrimination coefficient $r_{it} < .25$. The remaining five new items (39, 40, 43, 45, 46) discriminated well within the test.

## 4.1.5 Distractor Analysis

In theory, incorrect answers should be distributed equally among the distractors. Each item in this test has three or four possible answers: one target, and two or three distractors. With one target and three distractors, a subject has a 25% chance of getting a particular item right by guessing. When one distractor is more attractive than the other two, then the likelihood of getting an answer right by chance increases from 25% to 50%, and guessing is more effective. However, "creating three equally plausible distractor pictures is not always possible" (Gerken & Shady, 1996, pp. 137–138). Children's responses were recorded as a choice of Pictures 1, 2, 3, or 4 for each item, starting from upper left, upper right, lower left, lower right. For most of the analyses, these were later coded into dichotomous variables 0 (wrong) and 1 (right). Missing values were coded as "wrong" and assigned a 0 for the general data analysis. However, for the distractor analysis, missing values are handled as real missing values.

There are two main reasons for revising distractors: (1) where a distractor has not been chosen at all; or (2) where a distractor has been chosen too often – and may be more attractive than the correct answer (Lienert & Raatz, 1998). Lienert and Raatz (1998) suggest that the facility index as well as the discrimination index should be investigated for all distractors.

### 4.1.5.1 Facility Index for Distractors

Ideally, participants' incorrect choices should be distributed equally among all of the item's distractors. Since test items should show different levels of difficulty, the facility value of the distractors can be different for each item (Lienert & Raatz, 1998).

For the distractor analysis, the facility value for all distractors across all items was computed. In the first step, all items that were considered for revision from the item pool were marked to this effect (see Appendix I-2). Analyzing the facility value of the distractors identified the following categories of distractors that needed to be revised: (1) distractors that were not chosen at all; (2) distractors that were chosen more often than the target; and (3) distractors that showed an unbalanced response pattern.

Some of the items that were also considered for removal based on the item analysis also had problems with the distractors. The results are presented below together with the results of the discrimination index of the distractor analysis.

#### 4.1.5.2 Discrimination Index for Distractors

The discrimination coefficient $r_{it}$ for the distractors should show a negative correlation, which means they should be chosen more often by participants with lower scores than by participants with high scores (Lienert & Raatz, 1998). The distractors were recoded; for every item one distractor was assigned a value of 1, and all other distractors and the target were assigned a value of 0. Each recoded distractor was then correlated with the total test score using a Pearson product-moment correlation. The entire distractor analysis is displayed in Appendix I-2.

#### 4.1.5.3 Results of Distractor Analysis

An item's distractors should have a balanced facility value $p_i$ and a negative value for the discrimination coefficient $r_{it}$. The results of the distractor analysis contributed to making the decision whether certain items or distractors should be revised or removed from the item pool. Such a decision must be made carefully because of the small size of the subgroup of Deaf children of Deaf parents ($N = 34$). The results for all distractors, even for those items which would later be removed from the item pool based on the item analysis, will be presented here.

The majority of the items show good results in terms of facility value and discrimination index. Items where the distractors were not chosen equally were ultimately neither removed nor revised because of the small sample size. The majority of the distractors provided a negative correlation, thus fulfilling one of the criteria defined above.

In the first category are distractors (14.2[31], 31.1, 37.2, 41.1, 44.2) that showed a positive correlation in the discrimination index $r_{it}$ which means that they were chosen by children with high scores. This indicates that they are unclear or ambiguous (Lienert & Raatz, 1998), and also (based on the facility value $p_i$) that they were chosen more often than the target. All these items (14, 31, 37, 41, 44) were already considered for removal from the item

[31] The pictures (answers) of the items are numbered clockwise: (1) upper left; (2) upper right; (3) lower left; and (4) lower right. For example, the picture of Item 14 in the upper right is numbered as 14.2.

pool based on the item analysis. The Distractor 14.2 for Item 14 was chosen by 20 children (with the target only chosen by 8). Similar results can be found for other distractors: For example, the Distractor 31.1 was chosen by 16 children and the target only by 8; the Distractor 37.2 was chosen by 16 children compared to one correct choice of the target; the Distractor 41.1 was chosen by 25 children compared to two correct choices; and Distractor 44.2 was chosen by 24 children compared to four correct choices. These results raise the questions of whether those distractors that were chosen much more frequently than the target and which show a positive $r_{it}$ might represent the correct answer better than the original target? All of these items test spatial verb morphology, with the majority of the items (14, 31, 41, 44, but Item 37) representing the spatial concept of *in front* or *behind*, expressing spatial relationships between one referent (person, car, animal) and another (car, house, box). It could be argued for these items (Item 14: target *behind*; Distractor 14.2: *in front*; Item 31: target *in front*; Distractor 31.1: *next to*; Item 41: target *behind*; Distractor 41.1 *in front*; Item 44: target *in front*; Distractor 44.2 *behind*) that the chosen distractors depict the concept of *in front* or *behind* better than the original foils. Returning to Pilot 2 of the first test version with Deaf adults (see Chapter 3, Section 3.2.1), the results do not fully support this explanation. For Items 14 and 44, four of five Deaf adults chose the target; for Items 31 and 41, three chose the target and each of the two other Deaf adults chose a different distractor. All five Deaf adults chose the target for Item 37.

This is not striking evidence that the original foils represent the targets better than one of the distractors, but it is more probable – considering the reviewed acquisition studies on spatial relations (see Chapter 2, Section 2.5.1.7) – that children have not yet acquired the adult-like form. The chosen picture might represent an earlier stage in development. For subsequent analysis (homogeneity index), these items will be removed from the item pool. The inclusion of these items in a standardization study will be discussed in the next chapter.

The second category consists of items where one of the distractors was not chosen by any of the children. These are all items with four possible answers. Three of the items had been considered for removal based on the item analysis (Items 2, 35, 38). In Item 40, two out of four distractors were never chosen; this suggests a need to revise those two distractors. For the remaining items in this group, only one distractor was never chosen (Items

P3, 1, 4, 7, 8, 16, 17, 22, 23, 24). This suggests the option of either revising those distractors or removing them, leaving two distractors for those items.

In the third category are distractors that do not have a negative correlation (and were not chosen in preference to the target). These are the Distractors 20.4, 34.3, 36.4, and 42.4. Items 36 and 42 are to be discarded based on the item analysis, while the Distractors 20.4 and 34.3 need to be revised.

Having examined the research question regarding the quality of the distractors, now the research questions regarding the homogeneity of the test will be addressed.

### 4.1.6 Homogeneity of the Test

In theory, all items of a test should represent the trait to be tested equally well. In reality, items can never represent the same trait equally; they represent different facets of a trait through the test. A measure to address the extent of the overlap between the different facets of a trait is the homogeneity of a test (Fisseni, 2004). The homogeneity of a test is also related to the facility value and discrimination index. A higher scatter of the facility value across all items indicates lower inter-item correlation. When the items of a test have high inter-correlation and less scattering of the facility value, the test is more homogeneous, with the items representing different facets of the tested trait. A high inter-correlation of the items also indicates a high discrimination index.

There are different means of investigating the homogeneity of a given test, for example, by (1) inter-item correlation, and (2) factor analysis. It was decided not to use a factor analysis because factor analysis is generally used to investigate (multi)dimensionality of a test (Bortz & Döring, 2005) and not its homogeneity.

The homogeneity index *H* was investigated by applying an inter-item correlation (Bortz & Döring, 2005; Fisseni, 2004). The higher the investigated homogeneity index *H*, the more homogenous is the adapted DGS test: The higher the correlation of items with one another, the greater the probability that they represent the same construct. The homogeneity index *H* can be calculated for each individual item as well as for the entire test. Only the items that were kept in the item pool following the item analysis were included in the investigation of the homogeneity index *H*.

An inter-item correlation was calculated using the Pearson product-moment correlation. The homogeneity index *H* for each individual item was then calculated using the formula of calculating the sum of all correlations of an item (minus the correlation with self), divided by the number of items (minus the correlation with self). The homogeneity index for the entire test is calculated by the sum of all item indices *H* divided by the number of items (Bortz & Döring, 2005). Briggs and Cheek (1986) suggest that a range from .20 to .40 indicates acceptable homogeneity of a test. The result for the entire test is $H = .35$, thus showing a high degree of homogeneity across all items. The individual item homogeneity indices ranged from .20 to .48 (for individual Item *H* indices see Appendix I-3).

### 4.1.7 Evidence for Reliability

The internal consistency of a test is calculated by a statistical analysis such as Cronbach's alpha or split-half-analysis (Rust & Golombok, 2000). Other consistency measures, such as inter-rater reliability – comparing the scoring of a test by two raters – or test-retest reliability (Kline, 2000; Rust & Golombok, 2000) – correlating the scores from use of the same instrument on two occasions – are not appropriate for this study. Inter-rater reliability is not necessary since all test results were saved automatically on a computer hard disk. Test-retest reliability is not applicable, since scores were only obtained on one occasion.

Methods for investigating internal consistency are based on splitting a test into multiple comparable parts. With this approach, it is possible to measure the internal consistency of all items and not only of two comparable parts as with a split-half analysis (Lienert & Raatz, 1998). The method used most often is Cronbach's alpha. A special case of Cronbach's alpha for dichotomous coded variables is the Kuder-Richardson 20 formula (KR-20; Bühner, 2006; Lienert & Raatz, 1998). The pre-set Alpha-model used for calculating Cronbach's alpha can be applied using the KR-20 formula with SPSS (Bühner, 2006). The value of Cronbach's alpha increases with the number of items (Bortz & Döring, 2005). A minimum value of .70 can be considered as an acceptable value for a Cronbach's alpha (Nunnally, 1978).

The reliability coefficient Cronbach's alpha was calculated for the subgroup of Deaf children of Deaf parents on all items ($N = 49$) and then only on the items that remained in the item pool after the item analysis ($N$ = 39). Cronbach's alpha for all 49 items was $\alpha$ = .937. The Cronbach's alpha with the removed items based on the item analysis (39 items) increased to $\alpha$ = .955. The results confirm that the internal consistency of the adapted DGS test is high. Cronbach's alpha was also applied to all the linguistic sub-categories of the test (see Table 4.5) based on the remaining 39 items. The Cronbach's alpha for individual linguistic categories is lower – ranging from .470 to .896 – than Cronbach's alpha for the full set of items.

Table 4.5: Cronbach's Alpha for Linguistic Categories (Deaf Children of Deaf Parents; 39 Items)

| *Linguistic categories* | *Number of items in each category** | *Cronbach's alpha* |
|---|---|---|
| Handling classifiers | 3 | .470 |
| Negation | 11 | .829 |
| Number and distribution | 7 | .849 |
| SASS | 5 | .658 |
| Spatial verb morphology | 13 | .896 |
| Two-sign combinations | 2 | .638 |

* Two items (11, 32) occur in two categories

### 4.1.8 Evidence Based on Relationships with Other (External) Variables

In order to investigate whether the adapted DGS test shows a strong relationship to external but similar variables, the Spearman rank correlation coefficient $r_s$ (nonparametric) between the raw scores of the subgroup of Deaf children of Deaf parents ($N$ = 34) and the teachers' ratings of the children's receptive and productive DGS skills was performed. The teachers rated the expressive and receptive DGS skills of the children on a scale from 1 to 6 (1 indicating the highest level) represented on an ordinal scale. The data of the rating were recoded for the correlation (1 = 6, and 6 = 1). These ratings were available for 31 of the 34 Deaf children of Deaf parents. The correlations were performed with the original raw score, including the items to be discarded based on the item analysis.

The results reveal that there is a positive statistically significant correlation between the Deaf children's test performance and the teachers' ratings of their receptive DGS skills. The correlation ($r_s$ = .480, $p$ = .006) approaches that considered to be a strong correlation (.50) (Bortz, 1999; Cohen, 1992). These results mean that higher performance by the Deaf children on the adapted DGS test is correlated with higher ratings by the teachers of their receptive DGS skills.

A statistically significant positive correlation was also found between the Deaf children's test performance and the teachers' rating of their productive DGS skills ($r_s$ = .374, $p$ = .038). This is a medium effect (.30). The more important score is the correlation between teachers' ratings of receptive skills and the test scores, since receptive skills are the goal of the adapted DGS test.

These correlations need to be treated with caution, since the teachers themselves have different levels of signing skills. Teachers had been asked to rate their own DGS skills, for reception and production separately, in the educational background questionnaire (see Chapter 3, Section 3.3.2.3). They were provided with a scale describing five different skill levels (minimum score 1, maximum score 5). A translated version of the scale is provided in Table 4.6.

Table 4.6: Teachers' Scale for Self-Rating of Own Sign Language Skills

| *Sign language perception* | *Sign language production* |
|---|---|
| *Score 1:* I can comprehend several signs and simple sentences when they are signed slowly and with repetitions. | *Score 1:* I can produce a few signs (slowly) and reply to basic questions. |
| *Score 2:* I can comprehend basic/simple signed sentences, but I often have to ask for clarification in order to follow a conversation in sign language. | *Score 2:* I can produce basic sentences (but slowly), but I often have to think about how to express my thoughts/ideas in sign language. |
| *Score 3:* I feel quite confident in following a conversation in sign language, but occasionally I have to ask in order to understand everything. | *Score 3:* I feel quite confident in participating in a conversation in sign language, but occasionally I have to think about how to express my thoughts in signs. |
| *Score 4:* I can almost understand/follow all conversations in sign language. | *Score 4:* I can participate confidently in almost all conversations in sign language. |
| *Score 5:* I am able to comprehend conversations in sign language on any topic. | *Score 5:* I am able to participate actively in conversations in sign language on any topic. |

Thirty-nine teachers completed the questionnaire, but only 36 provided a self-judgment of their own DGS skills. Of these 36 teachers, 32 were hearing, and 4 Deaf or hard-of-hearing. The mean score for their self-rating of receptive skills for the hearing and Deaf teachers together was 3.28 (range = 1–5, *SD* = 1.29), and for productive skills 3.57 (range = 1–5, *SD* = 1.26). The mean for just the hearing teachers was slightly lower: 3.09 (range = 1–5, *SD* = 1.24) for receptive DGS skills; and 3.39 (range = 1–5, *SD* = 1.23) for productive DGS skills. The mean for the four Deaf teachers was higher: 4.75 (range = 4–5, *SD* = 0.5) for receptive DGS; and 5 for productive DGS skills. This difference between hearing and Deaf teachers is not surprising since these four teachers are more likely to use DGS as their preferred language.

The external variable of the teachers' rating provides supporting evidence for the validity of an external variable. However, these results should be treated with caution because the (hearing) teachers have different levels of DGS skills (range of scores = 1–5).

### 4.1.9 Evidence for Validity

Because no standardized DGS tests were available at the time of testing (2006), it was not possible to compare the results of an external measure with the test results of the adapted DGS test in order to investigate concurrent validity.

Only content validity was investigated. This was based on the review of linguistic studies of the structures represented in the adapted DGS test and sign language acquisition studies. These studies have already been presented in "Literature Review" and they will be discussed in the next chapter ("Discussion").

### 4.1.10 Test Performance of Deaf Children

This section addresses the issue of whether the test performance (for all items) is influenced by (1) gender of the children, (2) age of sign language exposure, (3) parental hearing status, and (4) chronological age.

For all subsequent statistical analyses, Fisher's exact test was computed to determine the significance of the relationship between dichotomous coded variables in two independent samples. Fisher's exact test is especially useful for small samples (Diehl & Staufenbiel, 2002).

Fisher's exact test (Table 4.7) was applied to see if there is a significant relationship between the raw score and (1) gender of children, (2) age of sign language exposure, (3) parents' hearing status, and (4) chronological age. For the purpose of applying Fisher's exact test, the variables needed to be recoded into categories, since this test cannot be used with continuous variables (age, raw scores). Therefore, the raw scores were categorized as low (0–25), medium (26–36), and high (37–49). This categorization was chosen in order to obtain a more or less equal number of children in each category. Five age groups were created: (1) 3;9 to 4;11, (2) 5;0 to 5;11, (3) 6;0 to 6;11, (4) 7;0 to 7,11, and (5) 8;0 to 10;10. Age of sign language exposure was coded as (1) birth to 3 years, and (2) 3;1 to 6 years. Parent hearing status was coded as (1) Deaf, and (2) hearing; and gender into (1) male, and (2) female.

Table 4.7: Fisher's Exact Test Across Raw Score and Gender, Age of Sign Language Exposure, Hearing Status, and Chronological Age

| | *Gender of children (N = 54)* | *Age of sign language exposure (N = 35)* | *Parent hearing status (N = 54)* | *Chronological age (N = 54)* |
|---|---|---|---|---|
| *Raw score (in categories)* | $p = .090$ | $p < .001$* | $p = .011$* | $p = .009$* |

*exact significance, 2-tailed

Results indicate a significant relationship between raw scores and the three measures of age of sign language exposure, parent hearing status, and chronological age ($\alpha = .05$). The nature of this relationship will be investigated in the following sections. A non-significant relationship was found between raw scores and gender, which was confirmed by a Mann-Whitney U Test ($U = 284.5$, $p = .175$).

#### 4.1.10.1 Evidence Relating Age of Sign Language Exposure to Test Performance

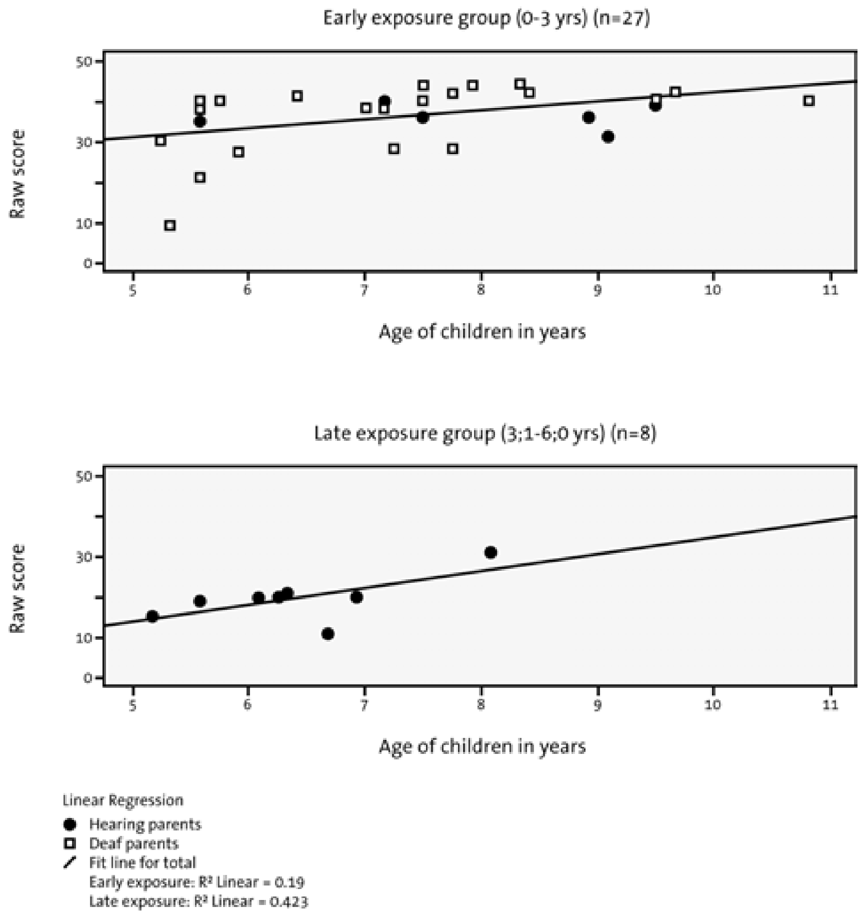
It can be assumed that all hearing children have access to a language from birth. However, for Deaf children, the situation is different since only 5% have Deaf parents (Mitchell & Karchmer, 2004) and, most likely, few have early exposure to a language from birth. This means that for the vast majority of Deaf children, language acquisition poses a considerable challenge (Marschark, 2002). Therefore, Deaf children constitute the only population group where timing of access to a language is a crucial variable. Late exposure to a L1 is a crucial variable in the subsequent mastery of this L1, as compared to the mastery of a (sign) language in children born Deaf who acquire a sign language from birth, or late deafened children who acquired English as their L1 and ASL as a late L2 (Mayberry et al., 2002). Therefore, the degree to which early L1 acquisition of DGS vs. late L1 acquisition of DGS accounts for variation in test performance is of interest.

Age of exposure information was only available for 35 of the 54 children (Table 4.8). The group of those with early exposure comprised 27 children (21 with Deaf parents, and 6 with hearing parents), whose mean age was 7;5 (range = 5;3–10;10). The group of children with late exposure comprised 8 children, all with hearing parents, whose mean age was 6;5 (range = 5;2–6;5).

Table 4.8: Parents' Hearing Status and Age of Sign Language Exposure ($N$ = 35)

| *Age of sign language exposure* | *Deaf parents (n)* | *Hearing parents (n)* | *Age range (M)* |
|---|---|---|---|
| Birth to 3 yrs (*n* = 27) | 21 | 6 | 5;3–10;10 (7;5) |
| 3–6 yrs old (*n* = 8) | 0 | 8 | 5;2–8;1 (6;5) |
| Total | 21 | 14 | |

Figure 4.2: Raw Scores of the Early and Late Exposure Group (Controlled for Parents' Hearing Status)



A univariate analysis of variance (ANOVA) was computed to compare the difference in test performance between the early and late exposure group[32]. The results reveal that the children with early exposure performed significantly better with a mean score of 36.04 than the late exposure group with a mean score of 19.63 ($F = 28.95$, $df = 1$, $p < .001$). The mean chronological age of the early exposure group ($M$ = 7;5) as compared to the late exposure group ($M$ = 6;5) is not statistically significant different ($F = 3.11$, $df = 1$, $p = .087$).

The scatter plot represented in Figure 4.2 provides additional information on the comparison between the two groups that differ on age of exposure to DGS. The late exposure group is much smaller ($n = 8$) than the early exposure group ($n = 27$), and the groups differ in terms of the percentage with Deaf parents 21/27 in the early exposure group; 0/8 in the late exposure group. Therefore, the results of this comparison are not independent of parents' hearing status since there is a substantial degree of overlap between the variables of (1) age of sign language exposure, and (2) parents' hearing status.

In a next step, a univariate analysis of covariance (ANCOVA) was calculated with the chronological age variable as the control variable (covariate) to see if chronological age also accounts for performance differences on the adapted DGS test. Controlled for chronological age, the main factor early vs. late exposure explains more of the performance differences between the two groups ($F = 23.42$, $df = 1$, $p < .001$) than the chronological age covariate ($F = 8.4$, $df = 1$, $p = .007$). Chronological age still has an influence on performance, but not as strong as age of exposure.

Next, it was attempted to adjust the chronological age variable for children in both groups by investigating the signing age[33] of the Deaf children. The signing age variable was investigated in order to see if the length of use

[32] At first, a nonparametric test for between-subject design, the Mann-Whitney U Test, was applied to compare the test performance of the early and the late exposure group. The early exposure ($n = 27$) group performed with a mean rank score of 21.48, which was statistically significantly better than the late exposure group ($n = 8$) with 6.25 ($U = 14$, $p < .001$). In a next step, a univariate ANOVA was applied with these two variables, with the raw score as the dependent and age of exposure as the independent variable. The results indicate that the early exposure group ($M = 36.04$) performed significantly higher than the late exposure group ($M = 19.63$) ($F = 28.95$, $df = 1$, $p < .001$) and thus confirmed the findings of the Mann-Whitney U Test. Therefore, an ANCOVA was applied to see if also other variables (e.g., chronological age, signing age) explain the difference in test performance between the early and late exposure group. There exist no nonparametric models in SPSS where control variables can be included.

of a sign language or the early vs. late exposure better explains differences in test performance. In her dissertation, Hoiting (2009) also investigated the effect of signing age when comparing the early lexical development of Sign Language of the Netherlands (NGT) in Deaf children of Deaf and hearing parents. The adjusted chronological age (i.e., signing age) of the group of Deaf children of hearing parents resulted in a similar learning trajectory to the Deaf children of Deaf parents, just at a different chronological age. In the study by Hoiting (2009), the goal was rather to show that Deaf children of hearing parents can catch up in their development when exposed to a sign language in the early ages (up to 3 years old). In contrast, in the present study the Deaf children of Deaf and hearing parents (for whom information on the age of exposure was available) are older (5;2–10;10), and therefore the adjusted age (i.e., signing age) of the Deaf children of hearing parents varies to a greater degree than the age at which they were first exposed to DGS. Including the variable of length of use of a sign language to control whether early vs. late exposure to a language in Deaf individuals is a crucial variable has been used in studies of Deaf adults with different ages of sign language exposure (e.g., Mayberry et al., 2002).

The information provided for age of exposure to DGS was only available for the 35 children in the early and late exposure groups. A descriptive overview (Table 4.9) of the different mean signing ages in both groups show that signing age is not independent of the variable of early and late exposure (as it was for hearing status), and that there is a certain overlap. Because of this difference in mean ages, it was not used as a covariate. Matching the signing age in the early and late exposure groups resulted in a reduced sample ($N$ = 25) of two strongly unbalanced subgroups (early exposure: $n$ = 24, late exposure: $n$ = 1) so that no further analyses were conducted to investigate the effect of signing age.

[33] The term *signing age* refers to the length of sign language use. For example, a 6-year-old Deaf child of Deaf signing parents might have a signing age of 6 years, whereas a 6-year-old Deaf child of hearing parents who started to have exposure to a sign language with 3-years of age has a signing age of 3 years (e.g., Hoiting, 2009).

Table 4.9: Chronological Age and Signing Age in the Early and Late Exposure Groups ($N$ = 35)

| *Exposure group* | *Deaf parents (n)* | *Hearing parents (n)* | *Chronological age range (M)* | *Signing age range (M)* |
|---|---|---|---|---|
| Early exposure ($n$ = 27) | 21 | 6 | 5;3–10;10 (7;5) | 5;3–10;10 (6;10) |
| Late exposure ($n$ = 8) | 0 | 8 | 5;2–8;1 (6;5) | 0;5–4;7 (1;9) |

In sum, the results suggest that early exposure to DGS has an impact on test performance. However, the results are only explaining a relation, not causation. The chronological age control variable also has an impact on test performance, but not as strong as the age of exposure factor. The parents' hearing status has an overlap with the two groups of early and late exposure to DGS and therefore the age of exposure variable is not independent of parents' hearing status. Also crucial is the different $N$ in both groups, 27 in the early but only 8 in the late exposure group. The signing age variable could not be used to adjust for chronological age in order to see if length of DGS use might account for a different test performance than the chronological age variable. Additionally, signing age is not independent of the early vs. late exposure factor.

#### 4.1.10.2 Evidence Based on the Hearing Status of the Parents and Raw Score

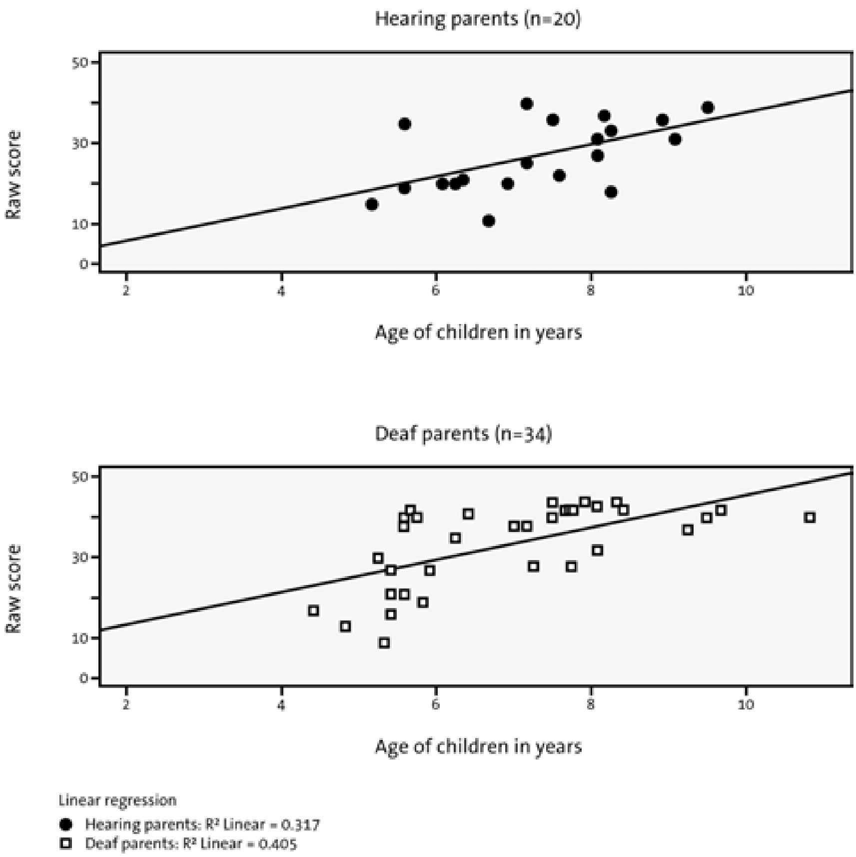
It has been reported in the literature that Deaf children with Deaf parents have better sign language skills than Deaf children with hearing parents (Strong & Prinz, 1997, 2000). However, the parents' hearing status per se is only one variable that explains the better sign language skills. The factors of early input at home and the age at which Deaf parents started to acquire a sign language might also make an important contribution to explaining DGS development (Singleton & Newport, 2004). Therefore – as in the previous section – it will be attempted to apply the signing age variable in order to see if it might additionally explain the different test performances of Deaf children of Deaf parents and of hearing parents.

Bearing in mind the different sizes of the subgroups of Deaf children with Deaf parents ($n$ = 34) and Deaf children with hearing parents ($n$ = 20), the scatter plot presented in Figure 4.3 provides a first impression of the different learning trajectories of both subgroups. The aim in describing both learning trajectories is to provide a general descriptive overview, without considering outliers. The Deaf children of hearing parents showed a more linear learning trajectory, with an increase in scores from 5 years onward, reaching the higher scores from 8 years onward (although the highest score of 40/49 was achieved by one child at 7;2). In contrast, the Deaf children of Deaf parents showed a sharp increase in trajectory between the ages 5–6 years old, reaching their maximal raw scores around 6–7 years old and then plateauing (the highest score of 44/49 was achieved by one child at 8;4). The trajectory of the Deaf children with hearing parents cannot be explained based on the data available for signing age (see below). However, studies of other sign languages (e.g., Brazilian Sign Language: Bernardino, 2005; NGT: Hoiting, 2009) suggest a delayed but still progressing trajectory in sign language development of Deaf children with hearing parents (depending on the age of exposure). The learning trajectory of the Deaf children of Deaf parents is likely to represent a normal DGS development. Either learning of DGS is completed by 6–7 years old, or later DGS development is not represented in the test items and therefore the test is insensitive to DGS development from 6 years onward, which is the more likely explanation. However, since the effect of signing age cannot be investigated, it is not entirely sure if these two learning trajectories vary empirically between Deaf children of Deaf and hearing parents.

Since the signing age variable could only be investigated descriptively (Table 4.10), the Mann-Whitney U Test was applied to investigate the difference in test performance between the Deaf children of Deaf parents and Deaf children of hearing parents. The distribution of raw scores between Deaf children of Deaf parents and Deaf children of hearing parents is statistically significant ($U = 197$, $p = .010$). The mean rank score for children with Deaf parents with 31.71 (age range = 3;9–10;10, *M* age = 6;10) was significantly higher than the mean rank score of 20.35 for children with hearing parents (age range = 5;2–9;6, *M* age = 7;4). The different mean ages (6;10 vs. 7;4) in both groups is not significantly different ($U = 268$, $p = .197$). Thus, there is evidence that parentage has a statistical significant impact on test performance, but it is not clear what causes the effect on better test performance.

In addition, a nonparametric correlation, the Spearman rank correlation coefficient, was performed to investigate the relationship between the hearing status and raw scores. The Spearman rank correlation coefficient ($r_s$ = .352, $p$ = .009) showed a statistically significant correlation with a medium effect size, providing additional support for evidence of a relation between parent hearing status and the children's raw scores: Higher raw scores are achieved by Deaf children of Deaf parents compared to Deaf children of hearing parents. Still, it does not explain what causes the effect on better test performance.

Figure 4.3: Raw Scores of Deaf Children of Hearing Parents and Deaf Children of Deaf Parents



In a next step, the chronological age variable was adjusted in the subgroups of Deaf children of Deaf and hearing parents. The signing age variable was only provided for 35/54 children. Matched for signing age in both subgroups, the sample size was reduced to 25 children. Therefore, it was decided to present the signing age variable only descriptively.

For that purpose, Deaf children of Deaf and hearing parents were categorized into three age bands in order to be able to match for age: (1) 4;0–4;11, (2) 5;0–5;11, and (3) 6;0–8;6. A total of 25 children (18 children of Deaf parents, 7 children of hearing parents) could be identified across the three age bands, with different numbers represented in each age band. In the first age group ($n$ = 4), the Deaf children of Deaf parents ($n$ = 2) showed a higher mean score, with 39 (range = 38–41), than Deaf children of hearing parents, with 33 (range = 31–35, $n$ = 2). In the second age group ($n$ = 6), five Deaf children of Deaf parents showed a mean score of 32 (range = 21–40) and one child of hearing parents showed a raw score of 36. In the third age group ($n$ = 15), the Deaf children of Deaf parents ($n$ = 11) revealed a higher mean score, with 39 (range = 28–44), than the Deaf children of hearing parents ($n$ = 4), with a mean score of 36 (range = 31–40). These descriptive data are summarized in Table 4.10.

The results of signing age revealed that Deaf children of hearing parents performed only slightly lower ($M$ raw score = 35) on the DGS test than the Deaf children of Deaf parents ($M$ raw score = 37). This was confirmed by a Mann-Whitney U Test ($U$ = 49.5, $p$ = .412). There is a non-significant difference between the mean rank scores of Deaf children of Deaf parents ($n$ = 18, $M$ age = 6;5, mean rank score = 13.75) as compared to the Deaf children of hearing parents ($n$ = 7, $M$ age = 6;4, mean rank score = 11.07). In sum, the data result in an $N$ which is too small to compare the effect of signing age on test performance between Deaf children of Deaf and hearing parents.

Briefly, parentage has an impact on the test performance, but it was not possible to investigate if that explains the better performance in Deaf children of Deaf parents (e.g., issue of early input).

Table 4.10: Descriptive Comparison of Signing Age and Raw Score in Both Subgroups ($N$ = 25)

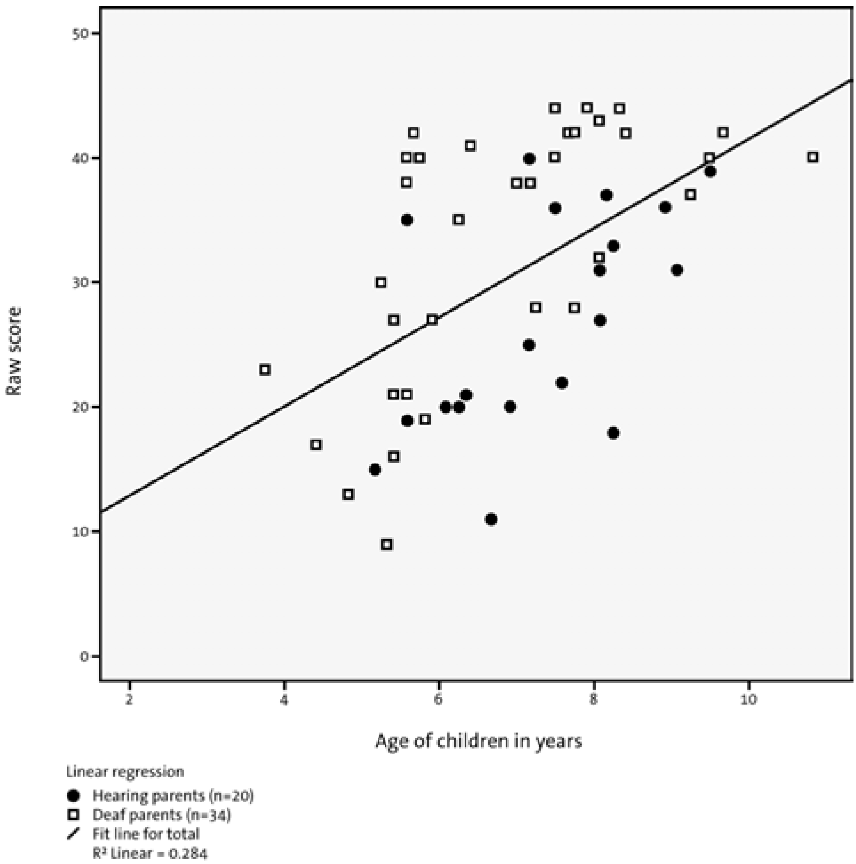
| *Parents' hearing status* | *Signing age bands and mean scores* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4;0–4;11 (*n*) | Mean score (range) | 5;0–5;11 (*n*) | Mean score (range) | 6;0–8;6 (*n*) | Mean score (range) | Mean score across age bands |
| Deaf parents (*n* = 18) | 2 | 39 (38–41) | 5 | 32 (21–40) | 11 | 39 (28–44) | 37 |
| Hearing parents (*n* = 7) | 2 | 33 (31–35) | 1 | 36 | 4 | 36 (31–40) | 35 |
| Total | 4 | 36 (31–41) | 6 | 32 (21–40) | 15 | 38 (28–44) | |

#### 4.1.10.3 Evidence Based on Chronological Age and Raw Scores

Adapting a test for sign language development requires that the adapted test be sensitive to age. In order to address the research question concerning a correlation between chronological age and raw scores, raw scores (ordinal scale/continuous variable) were correlated with child chronological age (continuous variable), for both the whole sample, and then separately for the subgroups of children with Deaf parents and hearing parents. For this research question, a nonparametric correlation method (Spearman rank correlation coefficient) was applied. The Spearman rank correlation coefficient $r_s$ between the chronological age and the raw score of the whole sample is significant ($r_s$= .530, $p$ < .001), indicating a strong effect size (.50). The correlation indicates that the older the child, the higher the raw scores (Figure 4.4).

For the subgroup of Deaf children of Deaf parents, the Spearman rank correlation coefficient indicates a strong effect size of the correlation ($r_s$ = .681, $p$ < .001) between chronological age and raw score. For Deaf children of hearing parents, the correlation was lower but still significant ($r_s$ = .541, $p$ = .014), indicating a strong effect size, although smaller than for Deaf children of Deaf parents.

Figure 4.4: Raw Scores and Chronological Age of Whole Sample (*N* = 54)




Additionally, the different linguistic categories (e.g., negation, spatial verb morphology) across all 49 items were correlated only in the subgroup of Deaf children of Deaf parents. The results are presented in Table 4.11, suggesting a medium to strong effect size of correlation (range = .400–.754) between the chronological age of the children and the raw score of the linguistic categories. In sum, the older the children are, the more accurately they perform on the different linguistic categories.

Table 4.11: Correlation of Chronological Age and Raw Score by Linguistic Category (Subgroup of Deaf Children of Deaf Parents; 49 Items)

| *Linguistic categories* | *Spearman rank correlation coefficient $r_s$* | *p* |
|---|---|---|
| Handling classifiers | .604 | < .001 |
| Negation | .621 | < .001 |
| Number and distribution | .754 | < .001 |
| SASS | .400 | .019 |
| Spatial verb morphology | .608 | < .001 |
| Two-sign combinations | .691 | < .001 |

In a next step, the Deaf children with Deaf and hearing parents together were grouped into three age bands and then correlated with the raw score. The results are displayed in Table 4.12. In the first two bands (3;9–5;11 and 6;0–7;11), the correlation between chronological age and the raw scores show a strong correlation, but in the first band, the Deaf children of Deaf families (15/18) outnumber the Deaf children of hearing families (3/18). The second band is more balanced (Deaf parents: 11/20; hearing parents: 9/20). In the third age band (8;0–10;10), the correlation is not significant, suggesting that from > 8 years old there is no relation between chronological age and raw scores. Thus, the items are not sensitive enough at the older ages. The number of Deaf children with Deaf and hearing parents was balanced in the third age band (8 children in each group).

Table 4.12: Correlation by Age Bands and Raw Score for the Whole Sample ($N$ = 54)

| *Age band (n)* | *Deaf parents (n)* | *Hearing parents (n)* | *Spearman rank correlation coefficient $r_s$* | *p* |
|---|---|---|---|---|
| 3;9–5;11 (18) | 15 | 3 | .532 | .023 |
| 6;0–7;11 (20) | 11 | 9 | .597 | .005 |
| 8;0–10;10 (16) | 8 | 8 | .394 | .131 |

These results provide important evidence relating to the research question of whether age correlates with the test scores.

*Descriptive comparison of items representing different spatial concepts*: It has been argued in this section that the chronological age variable is related to the raw scores of the Deaf children. This investigation was across all items, independent of which linguistic structure they represent. Based on the available data, it is not possible to make any statistical analyses of whether chronological age and/or signing age are related to the linguistic structures that were pointed out in "Literature Review" to be acquired at different ages (e.g., classifier constructions). But here, descriptive comparisons will be presented between items representing simple and more complex linguistic structures encoding different spatial concepts.

Different spatial concepts with varying degrees of complexity encoded in classifier constructions have been identified in Chapter 2, Section 2.5.1.7. For example, spatial concepts representing *on* or *in* are less complex and are acquired earlier than the more complex *right-left distinction* or *behind*, (e.g., Martin & Sera, 2006).

More complex items representing spatial concepts such as *in front*, *behind, top-right*, *below-left*, or *inside-left* have not been solved by many children (Item 14: 8/34; Item 31: 7/34; Item 36: 19/34; Item 37: 1/34; Item 41: 2/34; Item 44: 4/34; but Item 35: 29/34). All these seven items have been identified for removal (or revision) based on the item analysis (mostly because of the item discrimination; see item analysis, Appendix I-1). These spatial concepts are suggested to be acquired later (Martin & Sera, 2006; Morgan et al., 2008; Slobin et al., 2003).

Five items in the adapted DGS test represent the spatial concepts *in*, *on*, or *under* (Items 5, 8: *on*; Item 10: *in*; Items 16, 40: *under*). When dividing the entire 49 items of the adapted DGS test (3 practice items, 46 test items) into three bands with equal number of items ordered from the easiest to most difficult items (Band 1 and Band 2 with 16 items, Band 3 with 17 items), the two items representing *on* are among the easiest items in Band 1 (among the top 7 items, Item 5 solved by 29/34; Item 8: 30/34); followed by Item 10, representing *in*, in the last position of Band 2 (solved by 23/34); and finally the last two items representing *under* at the end of Band 2 and the beginning of Band 3 (Item 40: 23/34; Item 16: 22/34), indicating similar levels of difficulty.

In sum, the assumed more complex items representing *in front*, *behind*, and *right-left distinction* (which need to be revised) have not been solved by as many children as the items representing *on*, *in*, or *under*. Whether the results of more complex items, for example, *right-left distinction*, are indeed related to age of acquisition, as proposed in "Literature Review" (or if the performance might be influenced by other factors), cannot be investigated here statistically because of the small amount of data available (e.g., the more complex items have only been solved by very few children; see above). This will be further investigated in "Discussion", Chapter 5.

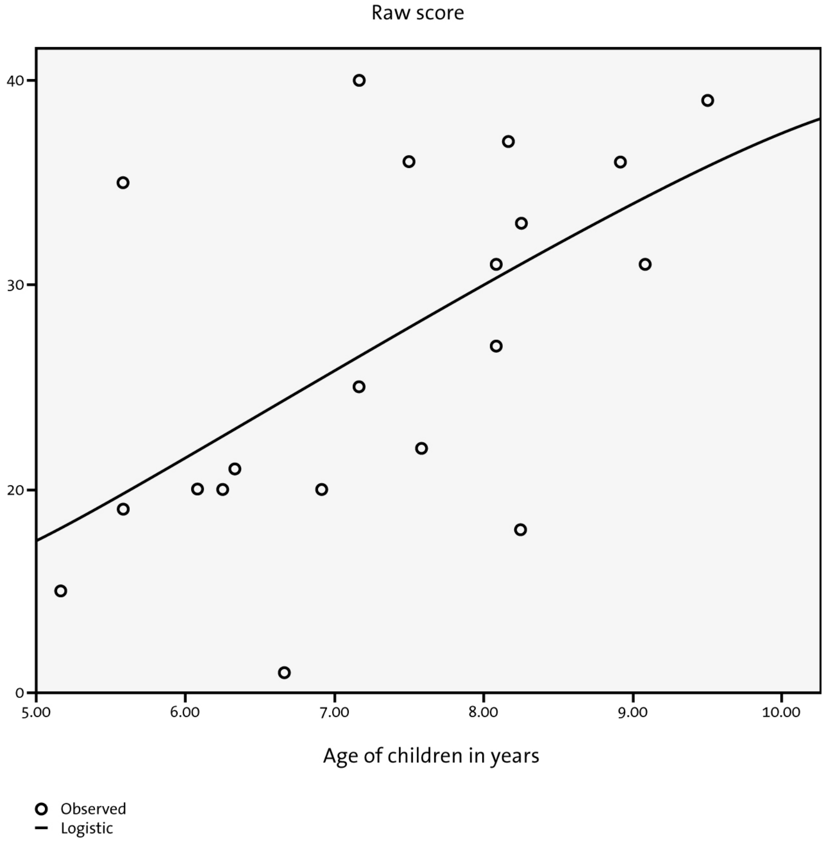
#### 4.1.10.4 Effect Based on Regression Models Between Chronological Age and Raw Scores

So far, only the research questions of whether there is a relation between the raw score on the one hand, and age of sign language exposure, parents' hearing status and chronological age on the other hand, have been discussed.

Because the age of exposure variable has a large overlap with the parents' hearing status variable, it will not be included in the regression analysis (see Chapter 4, Section 4.1.10.1). Since the two subgroups of Deaf children with Deaf parents and hearing parents show a different trajectory in their raw scores (see Chapter 4, Section 4.1.10.2), it was decided to perform independent regression analyses for each subgroup; each consisting of the dependent raw score variable and the independent chronological age variable.

A regression model provides the opportunity not only of seeing whether two variables are perfectly correlated, but also whether a set of variables, even when not perfectly correlated, can be used to explain the effect of one variable on another (Kiess, 1996). In the concrete case of the correlation of chronological age and raw scores, the main purpose is to investigate the effect of the chronological age variable has on the test performance in the two subgroups.

Figure 4.5: Regression Model with Logistic Curve Fit of Deaf Children of Hearing Parents (*N* = 20)




Even when both subgroups show different learning trajectories as previously presented descriptively in text format, a regression model with a logistic curve fit was applied because a upper bound value needs to be defined for the dependent variable (raw score) for that model (Janssen & Laatz, 2007). The upper bound value is 49 (maximum score). Additionally, a linear regression was calculated for both subgroups separately in order to see which of these models (logistic vs. linear curve fit) showed the better fit. Criteria deciding for the regression model with logistic curve fit were based on the values of $R^2$, standard error of estimate (*SEE*), and the significance of the independent chronological age variable (the results of the linear regression models of both subgroups are in Appendix I-4 and I-5). Additionally, a content-related argument is that a developmental curve in children is never linear since children have acquired language at a certain stage and the curve flattens out. Therefore, a regression model with logistic curve fit is the better fit for the purpose of investigating the effect of chronological age on the variable raw score.

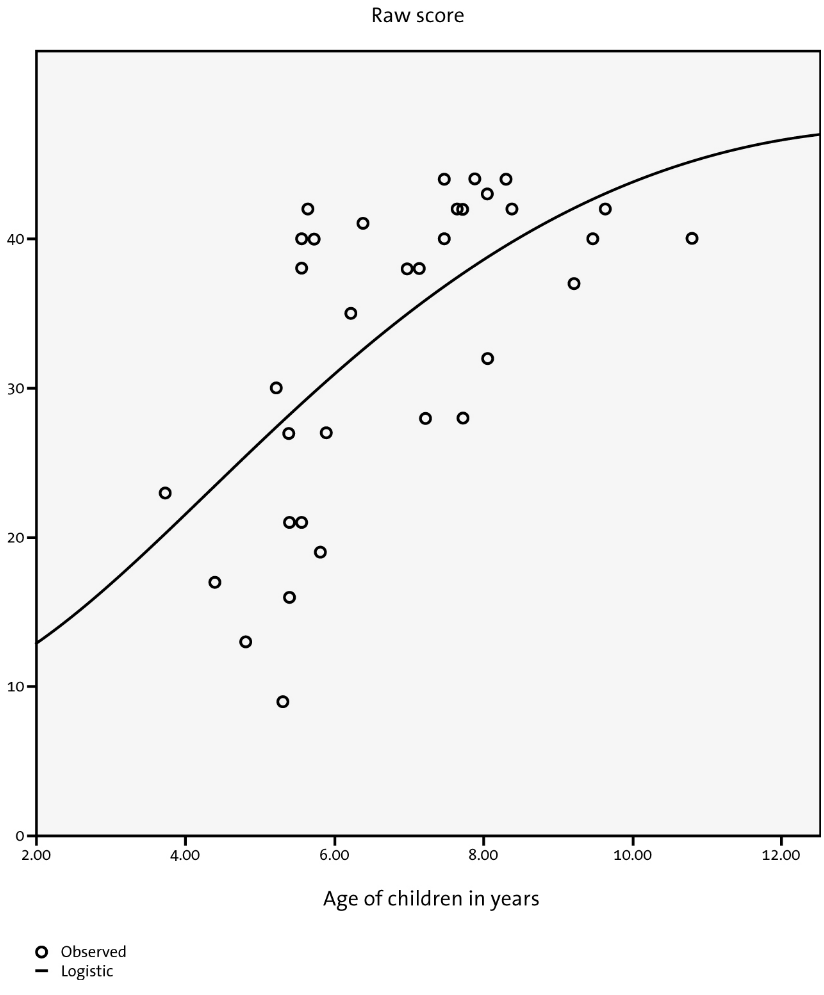
The results ($F$ [1, 18] = 7.94, $p < .001$, $R^2 = .306$, $SEE = 7.45$) reveal that in the subgroup of Deaf children with hearing parents, the chronological age factor is significant and that it has an effect on the raw scores within this subgroup. However, the $R^2$ value of .306 explains only 30.6% of the variance of the raw scores within this subgroup (Appendix I-6).

Figure 4.5 provides a scatter plot of the regression model with logistic curve fit for the subgroup of Deaf children of hearing parents.

For the Deaf children of Deaf parents subgroup a regression model with logistic curve fit was also applied. The results ($F$ [1, 32] = 21.17, $p < .001$, $R^2 = .398$, $SEE = .808$) reveal that in the subgroup of Deaf children with Deaf parents, the chronological age factor has a significant effect on the raw scores within this subgroup. However, the $R^2$ value of .398 explains only 39.8% of the variance of the raw scores within this subgroup (Appendix I-7).

Figure 4.6 provides a scatter plot of the regression model with logistic curve fit for the subgroup of Deaf children of Deaf parents.

Figure 4.6: Regression Model with Logistic Curve Fit of Deaf Children of Deaf Parents ($N$ = 34)



## 4.2 Summary

There is evidence for the effectiveness of the adapted DGS test. The results of the item and distractor analyses provide a good basis for the revisions of items in the test instrument required for successful standardization. The homogeneity index *H* provides support for the homogeneity of the tested construct of this test that is also important for further development. The Cronbach's alpha measure of internal consistency was found to be high across all items, and even increased when items were removed that were previously identified for removal based on the item analysis. The different linguistic categories of the test also showed – to different degrees – internal consistency.

An external variable, the teachers' rating of the children's DGS abilities, correlated significantly with the children's test performance, but the results have to be treated with caution considering the mean scores and the variability in the hearing teachers' own DGS skills.

The gender variable did not show any significant differences. The early exposure (0–3 years old) to DGS in Deaf children was found to have an impact on test performance, but it is not clear exactly what explains the differences with the late exposure group (3–6 years old). Also, the control chronological age variable revealed an influence on test performance, but not as strong as the age of exposure variable. The parents' hearing status also was supportive evidence for why test performance was higher in Deaf children of Deaf parents. However, adjusting for signing age did not work because not sufficient data were available. The results also indicated age sensitivity, for the whole sample as well as for both subgroups separately. No correlation could be found in children age > 8 years old. Age sensitivity is important for further test adaptation. The results of the regression model provided additional support for the age-sensitivity of the adapted DGS test.

Having presented the results from the data analysis, the next chapter addresses the discussion of these results and of the study as a whole.

# 5 Discussion

This chapter consists of six main sections: (1) evaluation of the adapted test instrument; (2) methodological considerations in relation to test adaptation; (3) adaptation model; followed by (4) limitations of the study; (5) directions for future research; and (6) summary and conclusion.

## 5.1 Evaluation of the Adapted DGS Receptive Skills Test

The aim of this study was to address cultural, linguistic, methodological, and theoretical issues in relation to the adaptation of a developmental language measure from BSL to DGS. In this section, the significant findings of this study will be discussed, starting with (1) cultural issues, (2) psychometric issues (item and distractor analysis, homogeneity index, reliability), (3) relations of raw scores to an external variable (teachers' rating of children's DGS skills), (4) other variables relating to performance differences (gender, language exposure, parents' hearing status, chronological age), (5) means of differentiation between test participants, (6) the variability of linguistic forms, (7) defining the norming sample, followed by (8) a finding that did not support the research questions.

### 5.1.1 Cultural Issues in Test Adaptation

Cultural issues were less dominant in the adaptation process to DGS than were, for example, linguistic and methodological issues. The major cultural issues, which have already been dealt with in Chapter 3, Section 3.1.1 refer mostly to culture-related concepts that are represented in the test materials of the source culture and that needed to be changed to fit in with the target culture (e.g., the differing British and German mailboxes). Similar issues have been reported for the adaptation of the BSL Receptive Skills Test to other sign languages (e.g., Danish Sign Language; Haug & Mann, 2008).

## 5.1.2 Effectiveness of Items

The findings of the empirically driven research questions provide a basis for item selection for a standardization study. (1) The findings taken together from the item analysis provided an indication as to which items should be discarded or retained in terms of their different levels of difficulty (item facility) and their ability to differentiate among children with high and low scores (item discrimination). This resulted in a suggested reordering of the items in the DGS Receptive Skills Test (Appendix J-1). (2) Some (5/10) of the newly developed items suggested good results and will therefore be retained in the item pool; three items were suggested for revisions. (3) The distractor analysis indicated that the distractors for some items would need to be revised (Appendix J-1). There was supporting evidence in connection with existing acquisition studies (e.g., Morgan et al., 2008) that some items should not be removed (as suggested by the results of the item analysis), but should rather be sampled again in a new pilot study. These items represent linguistic concepts that are important in language acquisition and should thus be included in the adapted DGS test. (For a complete overview of all these items, see Appendix J-1 and the discussion in Chapter 5, Section 5.1.7). (4) The results of the homogeneity index suggest that the test is homogeneous, representing the same trait across the items. (5) The results for reliability of Cronbach's alpha were good and showed strong internal consistency in the test ($\alpha = .955$).

### 5.1.2.1 Level of Difficulty of Items

The majority (29/49) of the adapted DGS items showed a facility value $p_i > .70$ (range = .706–.971), followed by 14 items ranging from $p_i = .529–.676$, leaving only 6 items with $p_i = .0.29–.324$. This suggests that in general the easier items ($p_i > .70$) outnumber the more difficult ones. The higher number of relatively easy items means that the test has a reduced possibility of differentiating between children with varying levels of DGS skills (see also Chapter 5, Section 5.1.7 below.)

### 5.1.2.2 Item Order

The item order (which reflects the items' level of difficulty) of the BSL test was compared to the item order of the adapted DGS test. Since no data on the facility value (and item discrimination) of the BSL test were available, the comparison is only descriptive (see also "Results of Item Analysis," Appendix I-1). The nine newly developed items and the three practice items were excluded from this comparison, but items that were suggested for removal or revision were not excluded. The 37 items were divided into four bands from the easiest (Band 1) to the most difficult (Band 4) (Bands 1, 2, and 4 had 9 items each, Band 3 had 10 items). Identical bands were created for the BSL test. In Band 1, 7/9 items and in Band 4, 6/9 items of the adapted DGS test appeared in the same item order band as in the BSL test. In Band 2, only 3/9 and in Band 3, 5/10 items appeared in both tests. Taking the two middle bands together, 15/19 of the items appeared in comparable bands in both the BSL and the adapted DGS test.

In sum, among the upper 9 and the lower 9 items a general trend of comparable item order between the BSL and the DGS test could be observed. Taking the two middle bands together, a similar pattern emerged. However, the original BSL item order was not maintained in the adapted DGS test. A possible explanation is that the adaptation of the BSL items to DGS worked well to a certain degree, but that language-specific structures might have caused certain BSL and DGS items to differ in their degree of difficulty. The source of similarities and differences in item order (representing level of difficulty) between the BSL and adapted DGS items cannot be investigated here since no data relating to the ordering of the BSL test items were available, and the item order of the BSL test is the result of a standardization whereas the adapted DGS test has not yet been standardized. This probably represents the pattern of general similarities in language development combined with language-specific differences.

### 5.1.3 External Variable: Teachers' Rating of Children's DGS Skills

Having a valid external measure of the Deaf children's signing skills at hand, which can be compared with their test performance, is important for the validation of a newly adapted test. Available for this study were the (Deaf and hearing) teachers' ratings of the Deaf children's DGS skills, which reveal what is approaching a strong correlation for comprehension ($r_s = .480$, $p = .006$) and a medium correlation for production ($r_s = .374$, $p = .038$). These correlations provide good evidence for an external source of validation. However, these data have to be treated with caution because the majority of the teachers (32/36) who provided the information on the children's DGS skills were hearing and had differing levels of DGS skills (see Chapter 4, Section 4.1.8).

Other studies have also addressed the issue of whether teachers' ratings of Deaf children's sign language competence provides a valid external measure of these children's signing skills. Herman and Roy (2006) found a correlation between testers' ratings ($N = 3$) prior to the test administration and Deaf children's scores on the BSL Receptive Skills Test. All three testers were experienced in working with Deaf children. It is not clear whether these three testers were Deaf and no information about their BSL skills has been provided, but it can be assumed that they had a good command of BSL since they functioned as test administrators. Herman and Roy (2006) consider that these results support the validity of the BSL test. In contradiction to this finding are the results of the BSL Receptive Skills Test adapted to Auslan (Johnston, 2004). Johnston found that the children's test scores did not match with the impressions of the teachers based on their everyday interaction with them. All Deaf and hearing teachers seemed to have had good Auslan skills (Johnston, 2004). Herman and Roy (2006) suggest that the results of the Johnston (2004) study might bring into question the validity of the adapted Auslan test. These results are not necessarily a result of an adaptation, but could also occur in the process of test development. Although it is not clear what caused the different results in the two studies (Herman & Roy, 2006; Johnston, 2004), one could say generally that in order to be able to use teachers' ratings of Deaf children's signing as a measure of validity of a newly adapted or developed test, it is important to have additional information on the teachers' own signing skills.

For future research, it would be advisable to revise and standardize the self-rating scale used in the present study in order to get a better measure of the teachers' DGS skills. The Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) might provide a model for the revision of the DGS self-rating scale[34]. If an effort is made to have ratings only from teachers who are known themselves to have a good command of DGS, it should be possible to match the teachers' own rating with

[34] The CEFR is a set of guidelines that describes the progress in learners of a foreign language across Europe. It aims to be used for evaluation and teaching. The CEFR refers to six different reference levels of language proficiency (A1 as the lowest level, followed by A2, B1, B2, C1, and C2 as the highest level) across different domains, such as understanding, speaking, and writing (Council of Europe, 2001).

the rating of the Deaf children. A standardized self-rating scale could also be used for Deaf and hearing people in other research studies, as well as in programs such as interpreter training where it is important to evaluate levels of DGS skills.

The variability of the self-rating among teachers also suggests that hearing teachers need more DGS training (Audeoud & Haug, 2008; Haug & Hintermair, 2003).

### 5.1.4 Content Validity

The research question relating to the evidence of content validity of the adapted DGS test was not empirically based, and will therefore be discussed here.

Content validity was defined in "Literature Review" as being present if, for example, the test items (and the test as a whole) represent the linguistic structures to be tested (Davies et al., 1999). This issue was approached by reviewing (1) studies on DGS structures represented in the BSL template (e.g., negation, spatial verb morphology), (2) studies that highlight cross-linguistic differences, and (3) studies on sign language acquisition. As for DGS studies and studies that highlight cross-linguistic differences, it has already been argued in the Chapter 2, Section 2.5.2 that comparable linguistic structures represented in the BSL test also exist in DGS, while there are also DGS-specific structures absent from BSL that need to be represented. It can therefore be argued – considering the state of research on DGS – that content validity based on a review of research literature can be found in the adapted DGS test. The studies on sign language acquisition provide evidence – although mostly from the acquisition of sign languages other than DGS – for the developmental aspect of specific linguistic structures that are represented in the adapted DGS test.

### 5.1.5 Other Variables Explaining Performance Differences

Other variables that are important for successful test adaptation involve factors which have been identified as potentially affecting scores (Herman & Roy, 2006; Johnston, 2004). These variables were presented in Chapter 4 and include (1) gender, (2) age of exposure, (3) parents' hearing status, and (4) chronological age.

(1) The adapted DGS test does not show any gender differences in the performance of male and female participants. The same lack of difference was reported for the original standardization study of the BSL Receptive Skills Test, but not for the later analysis of additional score sheets of the BSL test (Herman & Roy, 2006). The results in the present study may similarly be due to the small sample size.

(2) The variable of age of exposure to DGS, represented by an early exposure group (birth to 3 years old) and a later exposure group (3–6 years old), is important for the adaptation of the DGS test because the participants' different linguistic experiences might explain their different levels of performance.

Early exposure has an impact on test performance, but does not provide a full causal explanation because performance may be influenced by other variables such as chronological age and signing age. The chronological age variable also accounted for performance differences in the two groups, but the effect of the age of exposure variable was stronger. Again, the sample was very small ($N = 35$) and the age of exposure variable is also closely related to parental hearing status (21/27 of the early exposure group had also Deaf parents). The signing age variable (i.e., length of DGS use) could not be investigated because of the limited information available. In studies of the impact of early first language acquisition on language processing by Deaf adults, where length of exposure was controlled, early exposure was found to be a crucial variable for successful early first language acquisition (e.g., Mayberry et al., 2002). The signing age variable should be investigated in a standardization study.

(3) Parents' hearing status also provides information that might explain differences in test performance. The results of the study show that there is a significant relation between parents' hearing status and their children's raw scores (with Deaf children of Deaf parents outperforming Deaf children of hearing parents) but since it was not possible to investigate the effect of signing age, the source of the difference is not clear (e.g., early input). The mean age of the two groups did not differ significantly. Reasons that explain performance differences are, on the one hand, interesting in regard to issues of language development but, on the other hand, also constitute a factor to consider for the reference group in a standardization study.

The Deaf children of Deaf parents also reach their highest scores when they are between 6–7 years old, suggesting that the adapted DGS test is not sensitive enough for children > 7 years old. This would also suggest the need to develop more difficult items for older age groups.

(4) The adapted DGS test yields a strong correlation between chronological age and raw scores and thus can be considered sensitive to age. Correlations of different age groups with performance showed that there is no significant relation between children > 8 years old (Deaf children with Deaf and hearing parents together) and raw scores. This is additional evidence that the test is not sensitive enough for children > 7–8 years, not only, as discussed above where the comparison is made between the two subgroups of Deaf children, but also when both subgroups are taken together. Similar findings were found adapting the BSL test to American Sign Language (Enns & Zimmer, 2009). This is in contrast to the BSL test, which is standardized and differentiates between children from 3–11 years old (Herman et al., 1999).

### 5.1.6 The Reference and the Target Groups of Language Tests

Related to the different variables (age of exposure, parents' hearing status, age) that contribute to an explanation of performance differences on the adapted DGS test is the issue of the definition of the *reference* and the *target* groups for the standardization of the DGS Receptive Skills Test. Reference group here refers to the sample/group for a standardization study.

Compared to the situation for spoken language tests, in sign language test adaptation and development the intended target group/user group is in most cases not identical to the reference group of the standardization. Deaf children who do not have access to a sign language within the most critical early years of their lives (4–6 years old; e.g., Mayberry et al., 2002; Newport, 2002) are the main target group for sign language evaluation and intervention. The reference group, however, should be Deaf and hearing (near native) signing children from Deaf and hearing parents. These children provide a model against which the performance of children with other types of language exposure can be measured and standardization can be made (Herman, 2002; Herman et al., 1998). Included in the reference group for the standardization study of the BSL test (Herman et al., 1998) were Deaf children of Deaf parents, hearing children of Deaf parents (with a native signing background), and Deaf children of hearing parents from bilingual programs, with older Deaf siblings, or with hearing parents with very good BSL skills. Herman et al. (1998) compared the scores of the Deaf children of hearing parents with the scores of the other two groups. The results showed that the Deaf children of hearing families did not perform differently than the other two groups of children, except in the youngest age group (Herman et al., 1998). These results indicate that Deaf children of hearing parents, when they meet the above stated criteria of early language exposure, can be included in a standardization study in order to able to comprise as homogeneous and as large a group as possible for the standardization study. One could argue that only Deaf children (and maybe also hearing children) of Deaf parents should constitute the reference group, but parental deafness per se is not a guarantee of early exposure to a sign language; Deaf parents' own experience of early or late exposure to a sign language can also be an important variable (Singleton & Newport, 2004). For future research and standardization, it will be necessary to collect more information on the languages used in the child's home and environment, and the age of exposure to these different languages by the parents and other people who communicate with the child.

Along with the heterogeneity of the population of Deaf and hard-of-hearing children, the researcher must also consider the large number of children from diverse cultural and linguistic backgrounds within this population (Große, 2003, 2004; Mann, 2008), as well as the ever-increasing number of cochlear-implanted children. It will be important to collect data on the sign language development of these different subgroups. While it would also be desirable to develop different norms for the DGS test for these different subgroups, this would be quite difficult to carry out considering the small size of the subgroup populations. Nevertheless, it might be possible to build up language profiles to help better understand the language acquisition of these different subgroups, in order to provide the basis for a better evaluation and monitoring of their language acquisition processes.

This issue will later be linked to the estimated size of the sample of the standardization study (see Section 5.1.8).

### 5.1.7 Means of Differentiation Amongst Participants

Test items in the adapted DGS test should be able to differentiate among the groups of children; for example, between younger and older Deaf children and/or between Deaf children with different linguistic experiences/exposure (i.e., early vs. late exposure, diverse cultural and linguistic backgrounds). The long-term goal – as a result of standardization – is a norm-referenced test for DGS development, where the performance of a child is compared to that of his/her normative group (Brown, 2004; Brown & Hudson, 2002). There are two main issues which should be taken into account in this attempt to differentiate among groups of children in the DGS test adaptation: items with different levels of complexity, and items of different frequency.

*(1) Item complexity:* Items representing spatial concepts that have been suggested to need revision or new sampling (Appendix J-1) could be used in a future test version as a means of differentiating between younger and older children. These are seven items (Items 14, 31, 35, 36, 37, 41, and 44) representing different spatial concepts such as *in front*, *behind*, *top-right*, *below-left*, or *inside-left*, which are acquired relatively late (Morgan et al., 2008; Slobin et al., 2003). These items were not comprehended by many children in the present study; they were most likely too young to perform correctly on these complex items (the oldest child in this study was 10;10 years old, range = 3;9–10;10, $M$ = 7;0). Items representing easier spatial concepts *on*, *in*, or *under*, are correctly responded to by more children (see Chapter 4, Section, 4.1.10.2). It would nevertheless be advisable to include these items in a standardization study, especially if the age range of the participants is extended up to 12 years of age. Besides, it would be necessary prior to a standardization to develop and pilot more items that cover the age range from > 7 years onward.

Another issue of differentiation relates to the items representing negation. The negator sign NICHTS1 (nothing) was used in five of the ten negation items (Items 3, 8, 28, 33, 36) (the BSL test used a number of different BSL negation forms). The remaining five items express negation in other ways, for example, by non-manual negation of a verb. This raises the issue of whether these items with the negator NICHTS1 are redundant, measuring the same linguistic structure five times. The same issue was raised in the adaptation of the BSL test to LSF (Haug & Mann, 2008). Therefore, it is suggested, working with a panel of experts to review the items representing negation, to add alternative forms of manual negation prior to a standardization study, as these might contribute to differentiation among the children (the issue of adapting negation from the BSL test to other sign languages has also been addressed in the Chapter 2, Section 2.3.2).

*(2) High and low-frequency:* Relatively low-frequency structures in a language are also a means of differentiation between younger and older children. Research studies on the acquisition of English have found that high-frequency structures tend to be acquired before items and structures that are of low frequency in the language addressed by adults to children (Tomasello, 2003). The state of research on DGS (which is not unlike that of many other sign languages) does not yet provide sufficient empirical data relating to frequency, let alone to DGS acquisition. However, the new, large 15-year DGS Corpus-Lexicon Project at Hamburg University will be gathering such data. Therefore, this point may be less problematic for future test adaptation and development in DGS, although data from this corpus project cannot account for the acquisition of high- and low-frequency structures in DGS. The way in which low- and high-frequency structures in DGS are linked to the complexity of linguistic structures is also worth investigating, as is complexity in relation to age of acquisition, as has already been pointed out in "Literature Review" (Section 2.5.1.7) (e.g., Morgan et al., 2008).

### 5.1.8 Defining the Norming Sample

In order to provide a figure for a norming sample for the standardization study of the adapted DGS test, the literature of other sign language standardization studies has been reviewed.

As for the vocabulary measure consisting of three tasks for DGS, written, and spoken German (the Perlesko), Bizer and Karl (2002) state that their norming sample consisted of 112 Deaf and hard-of-hearing children from 3$^{rd}$ to 5$^{th}$ grade from seven schools in Germany and one school in Zurich. The age of their sample ranged from 7;11 to 13;3 ($M$ = 10;5) for the norming of the DGS and written German tasks, while the age range for the spoken German task at the time of the norming study was 8;0–13;4 ($M$ = 10;6). Since the Perlesko is aimed for use in schools at these grade levels, the norming sample needs to be representative of this population. The authors claim that with 112 children from 3$^{rd}$ to 5$^{th}$ grade about 25% of this specific population is covered, thus the norming sample is representative. It is not clear how they worked out that the population of Deaf and hard-of-hearing children from 3$^{rd}$ to 5$^{th}$ grade consists of about 448 children.

Herman et al. (1998) conducted a standardization of the BSL Receptive Skills Test with 138 Deaf and hearing children. The authors defined different criteria for including children in the norming sample according to their linguistic experiences (Deaf and hearing native signing children and selected Deaf children of hearing parents with good BSL skills). They already made a pre-selection based on these criteria that children had to meet in order to be included in the standardization study. With their pre-selection, Herman et al. (1998) defined only a part of the entire population. This also raises the issue what constitutes the population of Deaf children and is strongly linked to the discussion earlier in this chapter about the reference and target groups (Section 5.1.6).

The norming sample of Hermans et al. (2010) included 330 Deaf children of Deaf and hearing parents aged 4–12 years old from seven of the eight schools for the Deaf in the Netherlands. Hermans and his colleagues argue that the norming sample was too small to allow them to investigate norms for different subgroups based on hearing status. Norms were established based on age groups. Since they covered 7/8 of schools for the Deaf in the entire Netherlands for the norming study, it can be estimated that this sample might be representative for the population of Deaf children in the Netherlands. However, no figures are provided for the total number of Deaf children in all eight schools.

It can be seen from the reviewed studies above that different approaches have been taken in defining the norming sample for a standardization study, and this mirrors the difficulty of defining the heterogeneous population of Deaf children in general.

Estimating an exact number required for a norming sample of the adapted DGS test is hard, since too many variables would need to be defined determining what constitutes the entire population of hearing-impaired children in Germany (e.g., including, Deaf, hard-of-hearing children, children with and without a cochlear implant). Possible variables for consideration are age, parental hearing status, gender, signing age, and linguistic background, including DGS and any other languages. Defining all the variables of the entire population is beyond the scope of this study, but should be tackled in future research.

From the other end, some figures for the number of hearing-impaired children in Germany have been reported by Große (2003). Citing different sources, an estimated number of 10,000 to 12,700 hearing-impaired children are in schools serving children and young adolescents with a hearing impairment. The number of hearing-impaired children in early intervention programs is estimated at between 2,500 and 4,000. Even when these figures represent the population of hearing-impaired children, it is not possible here to derive the number of children that constitute the target group of this test since it is not possible to specify the number of children that attend a school for the Deaf where some form of signing is used (which would be a requirement for using a test that measures DGS development).

Therefore, estimating the number of children required to take part in a standardization of the adapted DGS test will be approached by using the experiences of the other empirical studies: (1) defining qualitative criteria in terms of the linguistic experiences of the children (even though it would be preferable to include Deaf children of Deaf parents only) based on the study by Herman et al. (1998); and (2) defining six age groups (3;0–3;11, 4;0–4;11, 5;0–5;11, 6;0–7:11, 8;0–9;11, and 10;0 plus) covering the age range 3–12 years old with at least 30 children in each group (i.e., at least 180 children). This constitutes the minimum of potential subjects to conduct a standardization, but it should be attempted with more children if they are available. Yearly intervals in the younger age (till age six) are important, since language development is more marked at these ages (Herman et al., 1998; Hermans et al., 2010). Herman et al. (1998) included between 10 and 33 children in each of their age groups. Norming the adapted DGS Receptive Skills Test on different subgroups of children (e.g., Deaf children of hearing parents, Deaf children with diverse cultural and linguistic backgrounds) should be kept in mind as a long-term goal.

### 5.1.9 A Finding that Did Not Support the Research Questions

Not enough data on age of exposure was available to allow children to be matched for signing age. This would have provided data for an additional important variable affecting differences in DGS performance.

## 5.2 Methodological Considerations in Adaptation of Sign Language Tests

In this section, issues that impact on methodology will be discussed in relation to future test adaptation and development. Herman, Holmes, and Woll (2008) provide a short practical overview for the adaptation of the BSL Receptive Skills Test to any sign language. The goal here is to summarize methodological considerations based upon the experience gained from the adaptation to DGS and from the reviewed literature. Four sections deal with linguistic issues, followed by one section on validity.

### 5.2.1 Variability of Linguistic Forms and Test Adaptation

The variability of linguistic forms is an important consideration for the future development of different test versions. In order to account for regional variation of lexical items used in the vocabulary check and the adapted DGS Receptive Skills Test, data in three regions were collected (Chapter 3, Section 3.1.2). Regional variations were identified, but it was not possible in this study to assign all of them to a specific region. No other representative research on regional form variation was available at the time of the DGS test adaptation, although this will become available through the DGS Corpus-Lexicon Project.

As more research on regional variation becomes available, it will be easier to account for it in a DGS test by designing different versions for regional variations of DGS. Another option, one that was used in this study, is to introduce a training session where the signer modeling the items of the test teaches the signs identified as regional variations. While this approach has certain advantages, the test administrator needs to check that (1) the child understands the signs during the training session, and (2) the number of signs being taught is not so large that the child is unable to remember them over the course of the test. In the adapted DGS test, the training session included four lexical signs. The use of a training session has an advantage over different versions of a DGS test when children themselves do not consistently use signs from a single dialect.

## 5.2.2 Methodological Issues in Linguistic Research

In the context of the current state of DGS research, (1) the use of different models/theories in linguistic research addressing the same structure (e.g., classifier constructions), and (2) acquisition studies that report differences in the developmental timetable of certain structures (even when they can be explained by different methodologies used) pose a problem in applying research from other sign languages for adaptation to the target language.

The first point is especially relevant for studies on classifier constructions and their acquisition, where using different models/theories to explain the same structure hinders cross-linguistic research (Schembri, 2003). Applying the same model across sign languages might make comparison across sign languages easier (Schembri, 2003), which also is an advantage for test adaptation.

As for the second point, in an acquisition study of Auslan (de Beuzeville, 2006), the researcher used the elicitation materials from an ASL acquisition study (Schick, 1987), but found that the Australian Deaf children mastered classifier constructions earlier than the American Deaf children. Mastery was defined in both studies in terms of adult-like performance. The difference can be best explained by looking at how the two researchers defined their criteria for what constituted adult-like performance. Compared to Schick (1987), de Beuzeville (2006) accepted a wider range and different options of a specific structure as signed by Deaf adults as adult-like performance, which could explain the reported earlier mastery of classifier constructions in both studies (L. de Beuzeville, personal communication, April 23, 2009). These different results, even when explainable by the use of different criteria for adult-like production, need to be considered carefully when using these studies for test adaptation ("Literature Review," Section 2.5.1.7).

This issue is also related to the acceptability of forms in a language. Too little research is available on how different structures are actually used in discourse, as compared to isolated phrases in the cited acquisition studies and the context of the adapted DGS test. Here, again, the data to be collected in the new Hamburg Corpus-Lexicon Project which links lexical items to videos of signed texts, will be useful.

### 5.2.3 Acquisition Studies: Language Production and Comprehension

For sign language test adaptation, knowing when a certain structure emerges or is mastered in sign language production provides a first approach to including this structure in items in a comprehension test; based on the notion that comprehension precedes production (Hirsh-Pasek & Golinkoff, 1996; Morgan & Woll, 2002b, 2003). The exact timing of the development of comprehension of this structure is not clear, but this would be important to know in order to determine which linguistic structures should be included in a comprehension test for a specific age group.

The studies of sign language acquisition in "Literature Review" (Section 2.6.2) highlighted the predominance of production studies. Very few studies actually address sign language comprehension (the exceptions being, for ASL: abstract loci and verb agreement: Bellugi et al., 1988; for BSL: complex AB verb constructions: Morgan et al., 2002; Morgan & Woll, 2002b). Even here, comprehension of complex AB verb constructions in BSL (Morgan et al., 2002; Morgan & Woll, 2002b) is reported as emerging when children are around 3 years old, but this may just be the consequence of the age of children in the study, which was from 3 years upwards.

In sum, very little is known about sign language comprehension; there exists no model for comprehension (Morgan & Woll, 2002b). Clearly, more research is needed on the development of comprehension in different sign languages (and on sign language comprehension in general). More research would provide an important basis for the adaptation of sign language tests evaluating comprehension. A promising new approach in child language research on spoken language is the *preferential looking paradigm* (Hirsh-Pasek & Golinkoff, 1991, 1996), which studies early (12–30 months) language comprehension. This approach may also prove valuable as a method for researching early language comprehension in signing children.

### 5.2.4 Language-Specific Structures

A methodological drawback in adapting a test from a better-documented source language (BSL) to a less documented target language (DGS) is the state of research available. Since sign languages are not alike (e.g., negation: Zeshan, 2006), it is important to include language-specific structures in test adaptation. This requires a prior thorough review of existing research literature on that language. Much research on DGS is unpublished, mostly undertaken by Bachelor's or Master's program students. Such studies – even when unpublished – should be included in any literature review.

### 5.2.5 Validity of the Test: Linguistic or Visual-Gestural Representations?

Johnston (2004) theorizes about the validity of the adapted BSL Receptive Skills Test to Auslan. He discusses the issue of the integration of sign language grammars and gestures that are used in sign and spoken languages to convey meaning (Liddell, 2003) to the creation of sign language tests. He claims that apart from language-specific lexical items, the adapted Auslan test, as well as the original BSL test, might not actually test morphology and syntax, but rather that some of the structures (especially classifier constructions and number/distribution) "may be part of general visual-gestural representation strategies" (Johnston, 2004, p. 75). In other words, both the BSL and the Auslan instrument may be testing general non-verbal cognitive skills for all children or strategies "common to all users of a signed language as *modality linked features*; or, perhaps common to users of all languages – signed or spoken" (Johnston, 2004, p. 76). He further emphasizes that "language-specific features of BSL/Auslan, or any other language, must be learned rather than be features of any language *in that modality*" (p. 76).

This issue should be kept in mind for the standardization of the adapted DGS test – for example by including a pilot study with non-signing hearing children or adults. Which aspects of sign languages are not language- or even modality-specific, is an important question, although one that cannot be answered today considering the current state of research, and should be kept in mind for future research in test adaptation and development.

In sum, having a systematized approach at hand will further improve future test adaptations. A model for test adaptation will be discussed in the next section.

## 5.3 A Proposed Model for the Adaptation of Sign Language Tests

### 5.3.1 Background

In this section, the findings of the adaptation of DGS Receptive Skills Test will be used to produce a proposed *model for sign language test adaptation*. It has been argued in "Literature Review" that it is appropriate to use the approach of adaptation (e.g., van de Vijver & Poortinga, 2005) to transfer a test from the source sign language (BSL) to the target sign language (DGS). The model includes a number of empirical and methodological steps, which will be summarized after a discussion of the construct definition (these steps are also summarized in Appendix J-2)

The major theoretical contribution of this model of test adaptation is to define the construct and propose a methodological approach for validating this construct (by an external source), since one of the weaknesses of sign language instruments is the absence of reported psychometric properties (e.g., Haug, 2008a).

### 5.3.2 Approaching Construct Definition

Even when it can be assumed that both the original BSL and the adapted DGS test evaluate the same underlying construct (i.e., language development), the construct needs to be defined (van de Vijver & Leung, 1997a).

A construct can be defined as "an ability or set of abilities that will be reflected in test performance, and about which inference can be made on the basis of test scores" (Davies et al., 1999, p. 31). Within the process of test development, it is important to decide clearly what a test – in this case, language ability – aims to measure (Bachman, 1990; Douglas, 2000). This can be accomplished "by determining what specific characteristics are relevant to the given construct" (Bachman, 1990, p. 41). Van Dyk and Weideman (2004) define the test construct as a "blueprint [that] defines the knowledge or abilities to be measured by a specific test" (p. 7).

In the next step it is important to consider how the construct can be defined. Bachman (1990) defines language ability within his framework of *communicative language ability* (CLA). Bachman's theoretical framework of

CLA consists of three main components. One of these three components is *language competence*[35]. Language competence can be classified into two components: *organizational* and *pragmatic competence*. The organizational competence includes aspects of language form, for example, grammar, whereas pragmatic competence refers to language use. Each of these, in turn, consists of several categories and subcategories. Within organizational competence one of these subcategories includes the relatively independent competencies of vocabulary knowledge, morphology, and syntax (Bachman, 1990). This framework provides a characterization of what (could) form the target of language testing.

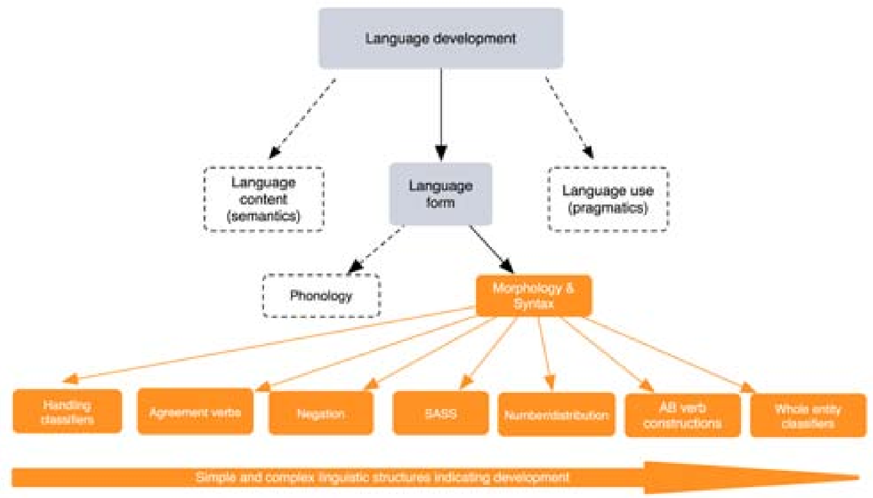
The construct of the proposed adaptation model will not be sited within Bachman's framework. Instead, Bachman's framework is used to derive methodological support to define the construct first on an abstract level and then further specify the several abilities that should be tested.

Tests of language development for children are concerned with linguistic content, *form*, and use; language form is concerned with the acquisition of word and sentence formation rules, that is, morphology and syntax (Wiig & Secord, forthcoming), which is also relevant for the adapted DGS test. In a similar way to how Bachman (1990) specifies different language competencies at different levels, it will be attempted here to approach the construct definition of the adapted DGS test based on the reviewed sign language studies.

The construct of the adapted DGS test can be defined as *development of morphology and syntax*. This includes simple and complex DGS structures of morphology and syntax, which relate to age and thus account for the developmental process. DGS morphology and syntax in the context of the adapted DGS test can be further specified to the different linguistic structures that should be represented in it (Figure 5.1). The order of the different linguistic structures (orange) as they are represented from left to right in Figure 5.1, do not indicate the exact sequence of their acquisition. However, the linguistic structures that are mastered first are further to the left, and those that are mastered later are to the right.

[35] Bachman (1990) uses the term *competence* not only limiting it to *linguistic competence* in the Chomskian sense, but also including, for example, pragmatic competence. Using the term *competence*, he refers to "entities which we may hypothesize to be in minds of language users" (p. 108). In turn, Bachman uses the term *ability* to include "knowledge or competence and the capability for implementing that competence in language use" (p. 108).

Figure 5.1: Components of Language Development Represented in Language Tests




The order of development of these different DGS structures can be determined by their varying levels of complexity across linguistic structures (e.g., handling classifiers vs. whole entity classifiers) and within different linguistic structures (e.g., whole entity classifiers referring to spatial concepts such as *on* or *in* are likely to be acquired before those referring to *in front*, *behind*, or *right-left distinctions*).

### 5.3.3 Operationalization of the Construct

The next step, and the most difficult part, is construct operationalization. Bachman and Palmer (1996) state that "the central activity of operationalization should be the development of test tasks" (p. 171). This step of defining the construct operationally "enables us to relate the constructs we have defined theoretically to our observations of behavior. This step involves, in essence, determining how to isolate the construct and make it observable" (Bachman, 1990, pp. 42–43), and hence construct validation. The operationalization of the construct of a sign language test involves adapting (or developing) the test items to the target language, accounting for (1) language-specific structures, and (2) developmental aspects in the target sign language.

One step in the operationalization of the construct is to rank the linguistic structures (represented in the items) according to their ascribed age of acquisition on the map of Ranking of Item Complexity by Deaf adults involved in the test adaptation process (see Figure 2.2 in "Literature Review"). This step will be referred to as *Ranking 1 of Item Complexity (operationalization).* The original item order of the BSL test can serve as a first indication for the ranking of the adapted items. The operationalization is followed by stages of piloting and revising of the adapted test items (see Figure 5.2 and Appendix J-2).

### 5.3.4 Validation of the Construct

Validation of a construct always aims at investigating what the test actually measures, that is, "construct validity concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or construct" (Bachman, 1990, pp. 254–255).

Following the operationalization of the construct and different stages of piloting and revisions, a validation of the construct by an external source is needed. One possible approach is letting Deaf experts who have not been involved into the adaptation process rank the different linguistic structures by their ascribed age of acquisition, and then compare this with the original ranking (Ranking 1 of Item Complexity; see Section 5.3.3) during the operationalization stage. This step in validating the construct will be referred to as *Ranking 2 of Item Complexity (validation)*. The original item order of the BSL test also provides a first indication of the level of difficulty for the adapted, although not for the newly developed, items of the DGS test. The results of Ranking 2 of Item Complexity can also be compared with the results of the item analysis, providing a first indication of the level of difficulty. This approach can be used not only for test adaptation, but also for test development.

A recent study where this ranking approach was used successfully (Vinson, Cormier, Denmark, Schembri, Vigliocco, 2008) lends methodological support to this approach to validation. Deaf adults ranked age of acquisition, familiarity, and iconicity of 300 lexical BSL signs in order to obtain norming data for these signs. The rankings were compared to a study of early lexical development of BSL which used parental checklists (Woolfe et al., 2010). The results revealed "a strong degree of correspondence between adults' estimates of the relative ages at which they acquired a sign and parents' judgments of what their children can actually comprehend and produce at a given age" (Vinson et al., 2008, p. 1085). It should be noted, however, that Vinson et al.'s BSL study rated lexical items only; in the case of the present study, the items to be ranked represent different levels of morpho-syntactic complexity. This has to be kept in mind for the standardization study since ranking morpho-syntactic items may be more difficult than ranking lexical items.

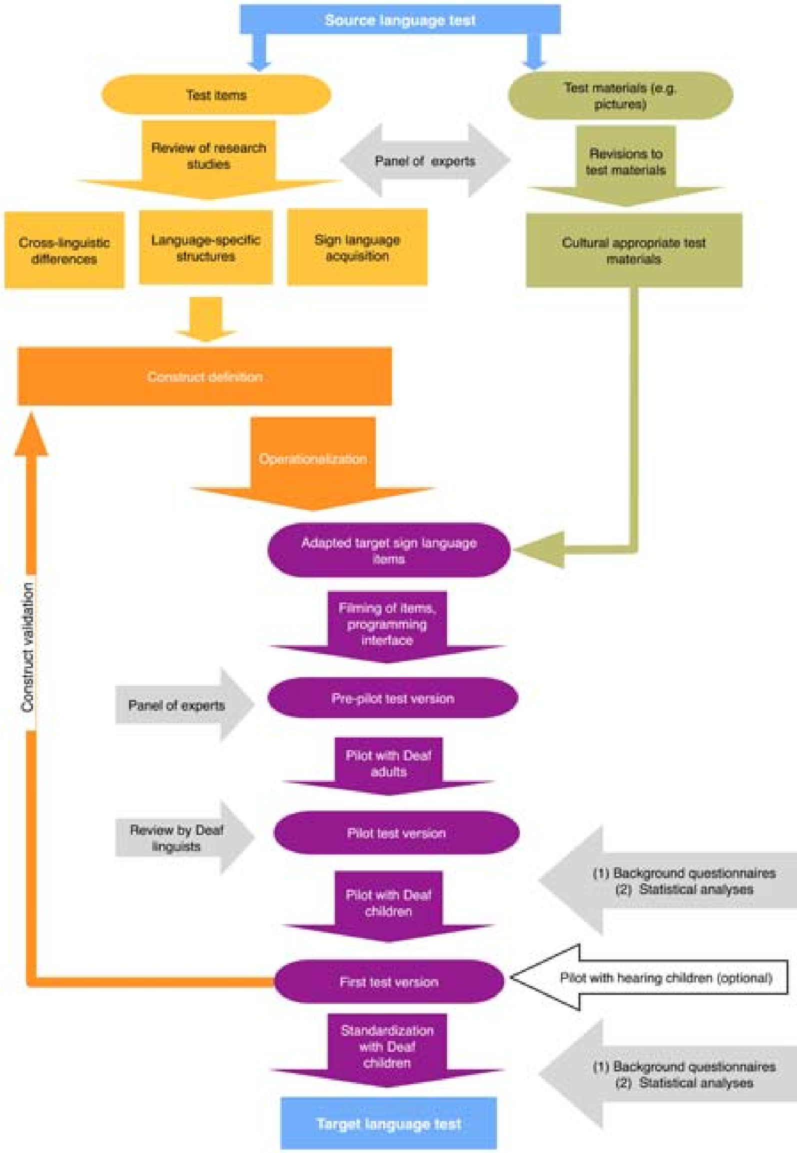
### 5.3.5 Proposed Model

It has already been argued that adaptation is the best approach to transferring a test across sign languages. The empirical steps involved in test adaptation have been described in the "Methodology" chapter, presented in the "Results" chapter, and discussed in this chapter. (The different empirical steps are displayed visually in Figure 5.2 and in Appendix J-2).

The main contribution of this model is to account for the construct definition and suggest a method for validating the construct of the adapted instrument in the light of the state of research in this field. In sum, adaptation of sign language tests must account for cross-cultural and cross-linguistic variation, developmental aspects, and establishing validity and reliability. The adaptation process is depicted in its entirety in Figure 5.2.

Boxes in light blue indicate the source and the target sign language tests. The subdivision below the source language test box into the light orange strand of *test items* and the olive green strand of *test materials* indicates the first steps in the adaptation process. The grey arrow with *panel of experts* provides input to both processes. The light orange strand includes the *review of the research literature*, the olive green strand *revisions of test materials* (any changes to test pictures). The light orange strand *review of the literature* provides empirical support/input for specifying the *construct* (orange strand) on different levels, which need to be *operationalized* and which then results in the set of adapted test items, which also receive input from the revised test materials.

Figure 5.2: Model of Sign Language Test Adaptation



The purple strand in the center of the model describes the different stages of *piloting and revisions* of the adapted test, which result in a standardized *target language test* at the end. Input from a panel of experts and linguists (grey arrows on the left side of the middle strand) will also be provided here. The grey arrows on the right side of the middle strand indicate that information will also be gathered using *questionnaires* and *statistical procedures* for the data analysis. The orange strand alongside the first test version indicates the process of *validation of the construct.*

Within the frame of the present study, the research only achieved the first test version stage preceding validation of the construct.

## 5.4 Limitations of the Study and Critical Self-Reflection

Although the study succeeds in answering (most) of the questions posed at the outset, there are some limitations that need to be addressed, and these will be discussed in the following sections.

### 5.4.1 The Target Group – Sample Size

One of the limitations of this study was the small sample size ($N = 54$). Obtaining large enough samples in statistical terms within this population of signing Deaf children in Germany is difficult to achieve because of its small $N$. The application of more complex statistical procedures to account for different variables (e.g., chronological age, age of exposure, parents' hearing status) in one model could not be carried out because of the sample size. Nevertheless, considering the stage of this research, as well as the one-to-one testing situation used, the findings of this study provided valuable insights on which to base a further revision of the test.

### 5.4.2 The Test Instrument

The computer-based methodology used here proved to have only minor limitations. One possible change might be to re-film the items to eliminate the circling movement made by the model at the end of each item to indicate that the children should pick one of the pictures. There was informal feedback that this movement is rather distracting[36].

### 5.4.3 Educational Background Questionnaires

There were some limitations to the questionnaires used in this study.

(1) To obtain valid information from the parent questionnaire in future studies, it is recommended that a translation of the questionnaires be provided to parents in at least some minority languages, for example, DGS and Turkish.

(2) As for the student's questionnaire that was filled out by the teachers, the scale of rating of the Deaf children's DGS skills by the teachers should be changed to a scale ranging from 1–10 in order to avoid confusion with the German grading system where 1 represents the best and 6 the lowest value. Also, an attempt should be made to get ratings of children's signing only from teachers who themselves have good DGS signing skills (e.g., Deaf teachers).

(3) Regarding the teacher's questionnaire, the self-rating scale should be standardized.

(4) For Pilot Study 2 with Deaf adults, it would be advisable to use an open-ended questionnaire to obtain more qualitative input from the Deaf adults, thus potentially yielding more data that might be valuable in test revisions.

### 5.4.4 Testing of Younger Children

Testing with the younger children showed that a computer mouse cannot always be used. Piloting with a child-friendly mouse or transferring the test to a computer which uses touch-screen technology should be considered.

### 5.4.5 Effect of Signing Age

Because of the limited data available, it was not possible to investigate the signing age on test performance. Furthermore, although questions about

[36]Selected examples of the DGS Receptive Skills Test can be accessed in the Internet at http://www.signlang-assessment.info/index.php/german-sign-language-receptive-skills-test. html.

age of exposure were included in the parent questionnaire, the answers could not be used as most parents did not understand this question. The provision of versions of this questionnaire in different languages should make more information available.

### 5.4.6 Validity

In order to increase validity, it was planned to have the adapted DGS test reviewed by a Deaf linguist, but resources were insufficient to permit this. This step should be included in the standardization study. The published DGS vocabulary test (Perlesko; Bizer & Karl, 2002) should also be used comparing the results of the Perlesko with the adapted DGS test to establish concurrent validity.

## 5.5 Directions for Further Research

Based on results of this test adaptation study, the following suggestions can be made:

(1) A standardization study should be undertaken, using the results, experiences and suggestions from this study. On the technological side, a web-based testing format could be used for this standardization, which would allow for more reliability in terms of scoring, etc. and larger numbers of participants. A pilot of a web-based testing format of the DGS Receptive Skills Test is under development.
(2) To be able to develop and adapt tests for DGS, more acquisition studies of DGS comprehension and production are needed,
(3) as well as more cross-linguistic sign language research in order to get a clearer understanding of the differences and similarities between sign languages and their acquisition.
(4) Also of importance is more research on variability and acceptability of linguistic structures beyond the lexical level (i.e., too little is known about morpho-syntactic structures and how they are used at the sentence and discourse levels).
(5) As previously pointed out, the self-rating scale of DGS for Deaf and hearing people should be standardized so that it can be used in different contexts.

## 5.6 Summary and Conclusion

Among the main contributions of the present study are new insights into the cultural, linguistic, methodological, and theoretical considerations necessary for future sign language test adaptation. The inter-connected cultural, linguistic and methodological issues were addressed at different stages of test adaptation; theoretical issues were addressed in the stage of construct definition. These steps have resulted in a proposed model for future test adaptation, covering empirical, methodological, and theoretical issues. This model of test adaptation can be applied to language test development for other under-documented sign languages.

On a more concrete level, the results and discussion of this study indicate further steps to be taken for the standardization of the adapted DGS test. The use of computer-based test technology with young Deaf children aged 4–8 years is a new and promising approach for future test adaptation and development.

The critically reviewed research studies offer a rich ground for discussing cross-linguistic sign language acquisition and thus contribute to a better understanding of similarities and differences across sign languages. On the one hand, the increasing number of sign language acquisition studies can serve to inform test adaptation and development, but on the other hand, data (expressive, receptive) provided by a larger number of Deaf children during test adaptation and development can contribute to a better understanding of sign language acquisition. The gain of knowledge is reciprocal for sign language acquisition and test development, and will contribute to further development in this field.

# 6 German Summary

## Evaluation und Adaption eines Verständnistests zur Deutschen Gebärdensprache

### 6.1 Problemstellung

Gehörlose Menschen benutzen häufig eine Gebärdensprache und die Mehrheitssprache in der gesprochenen und geschriebenen Form in ihrem alltäglichen Leben, das heißt sie können als bilingual angesehen werden (Grosjean, 2008). Der frühe oder späte Zugang zu einer Sprache hat einen Einfluss auf den Erstspracherwerb und als Konsequenz kann die Kompetenz in beiden Sprachen sehr unterschiedlich ausfallen (Mayberry und Lock, 2003). Nur 5 % der gehörlosen Kinder haben gehörlose Eltern (Mitchell und Karchmer, 2004) und erlernen so auch eine Gebärdensprache als Erstsprache. Für die restlichen 95 % stellt der Erwerb einer Sprache eine Herausforderung dar (Marschark, 2002). Der frühe Zugang zu einer Sprache und damit die Möglichkeit eines normalen Spracherwerbs zeigen auch längerfristige Vorteile in anderen Bereichen der Entwicklung von Kindern (Woll, 1998).

Die Gehörlosenpädagogik hat sich in den letzten Jahrzehnten verändert; vermehrt wird die bilinguale Erziehung gehörloser Kinder weltweit gefördert (zum Beispiel USA: Mahshie, 1995; Nover, 2005; Deutschland: Günther, 1999; Günther und Schäfke, 2004; Österreich: Krausneker, 2004; Dänemark: Lewis, 1995). Auch in Deutschland gibt es mehr Akzeptanz für bilinguale Versuchsklassen (zum Beispiel Hamburg: 1992; Berlin: 2001), auch wenn diese noch eine Minderheit innerhalb der pädagogischen Ansätze darstellen (Günther, Hennies und Hintermair, 2009).

Plaza-Pust und Morales-López (2008) evaluierten im internationalen Vergleich bilinguale Ansätze in Gehörlosenschulen und kommen zu dem Fazit, dass unter anderem fehlende Materialien und Testverfahren (zur Überprüfung der Gebärdensprachkompetenz) möglicherweise einen negativen Effekt auf die Auswertung von solchen bilingualen Konzepten haben. Dieser Bedarf nach Testverfahren für Gebärdensprachen wurde speziell in einer

Umfrage an Institutionen für Hörgeschädigte in Deutschland (Haug und Hintermair, 2003), aber auch international bestätigt (Schweiz: Audeoud und Haug, 2008; GB: Herman, 1998; USA: Mann und Prinz, 2006). Ein grundlegendes Problem ist das Fehlen von reliablen und validen Testverfahren, so dass Lehrpersonen häufig zu „selbst gestrickten" Verfahren greifen müssen, um eine sprachliche Bestandsaufnahme durchführen zu können (Singleton und Supalla, 2003).

Im Vergleich zu besser erforschten Gebärdensprachen wie der Amerikanischen Gebärdensprache (ASL: American Sign Language) oder der Britischen Gebärdensprache (BSL: British Sign Language) ist es hingegen für die DGS (Deutsche Gebärdensprache) schwieriger, einen Sprachentwicklungstest zu entwickeln, da es kaum Studien über den Erwerb von DGS gibt (Ausnahme: Hänel, 2003).

Eine der Ausnahmen von reliablen und validen Testverfahren zu Gebärdensprachen stellen der BSL Receptive Skills Test (BSL RST; Herman et al., 1999), der BSL Narrative Production Test (Herman et al., 2004) und der Perlesko, ein semantisch-lexikalisches Prüfverfahren zur Deutschen Gebärdensprache (Bizer und Karl, 2002), dar. Der Perlesko findet seinen Einsatz bei Kindern in der 3. bis 5. Klasse zur Überprüfung der semantisch-lexikalischen Kompetenz in deutscher Laut- und Schriftsprache sowie DGS. Es gibt kein Testverfahren in der DGS zum Testen der Verständniskompetenz von 4- bis 8-jährigen Kindern im Bereich Morphologie und Syntax.

Die wenigen vorhandenen Studien im Zusammenhang mit Gebärdensprachtests benutzen zum einen Gebärdensprachtests, um beispielsweise den Zusammenhang einer Gebärdensprache als Erstsprache und deren Einfluss auf den Erwerb der Mehrheitssprache in geschriebener/gesprochener Form als Zweitsprache zu untersuchen (zum Beispiel USA: Hoffmeister, 2000; Strong und Prinz, 2000; Deutschland: Mann, 2008; Schweiz: Niederberger, 2008). Zum anderen gibt es Studien, in denen die Entwicklung eines neues Tests zur Benutzung in den Schulen im Vordergrund steht (BSL: Herman, 2002; Niederländische Gebärdensprache [NGT]: Hermans et al., 2010).

Darüber hinaus gibt es Artikel, die sich auch auf das Thema der Adaption von Gebärdensprachtests beziehen (Johnston, 2004; Schembri et al., 2002), und einen Übersichtsartikel (Haug und Mann, 2008) zu diesem Thema. Aber es gibt keine empirischen Studien, die das Thema der Testadaption unter Berücksichtung kultureller, linguistischer, methodischer und theoretischer Gesichtspunkte, ausgehend von einem standardisierten Gebärdensprachtest, in der DGS untersuchen. Aufgrund der Forschungslage über die Struktur und den Erwerb von DGS ist die Adaption von einem vorhandenen und standardisierten Gebärdensprachtest ein „realisierbarer" Ansatz, der leichter in die Tat umzusetzen ist, als bei Null anfangen zu müssen. Die Lücke, die durch das Fehlen empirischer Studien entstanden ist, zu schließen, ist das Ziel dieser Studie. Des Weiteren wird unter Berücksichtigung der oben erwähnten Gesichtspunkte ein Modell zur Testadaption vorgeschlagen.

Das wichtigste Ziel dieser Studie ist es, Grundlagen für einen DGS-Test zu schaffen, der später nach der Standardisierung in den Schulen genutzt werden kann. Vom methodischen Gesichtspunkt aus betrachtet geht es um die Nutzung eines computerbasierten Tests. Aus theoretischer Sicht bietet diese Studie einen Beitrag zu kulturellen, linguistischen und methodischen Themen in der Testadaption. Als Adaptionsgrundlage dient der oben erwähnte BSL RST (Herman et al., 1999). Ein wichtiger theoretischer Beitrag zu einer erfolgreichen Testadaption ist das hypothesengeleitete Arbeiten, basierend auf Erwerbsstudien anderer Gebärdensprachen und Studien über DGS und andere Gebärdensprachen.

Im Folgenden werden die Forschungsfragen dargestellt:

1 Weist der adaptierte DGS-Test gute psychometrische Eigenschaften auf?

  Dies bedeutet im Detail folgende Unterfragen:

  1.1 Weist der adaptierte DGS-Test die Eigenschaften von Trennschärfe und Schwierigkeitsgrad auf?
  1.2 Wie passen die neu entwickelten Items in den gesamten Test?
  1.3 Weist die Distraktorenanalyse auf eine hohe Effektivität der Distraktoren hin?
  1.4 Gibt es Belege zur Homogenität des Tests?
  1.5 Gibt es Belege für die Reliabilität des Tests?
  1.6 Gibt es Belege für einen Zusammenhang zwischen einer externen Variablen, zum Beispiel der Einschätzung der DGS-Kompetenz der Kinder durch die Lehrer und den Rohwerten?
  1.7 Gibt es Belege für eine inhaltliche Validierung des adaptierten DGS-Tests? (Diese theoretisch geleitete Forschungsfrage wird in der *Diskussion* dargestellt werden).

2 Besteht ein Zusammenhang zwischen den Rohwerten, die die Kinder erreicht haben und anderen Variablen, wie (2.1) Geschlecht, (2.2) Alter des Kindes beim Zugang zu einer Gebärdensprache, (2.3) Hörstatus der Eltern und (2.4) Lebensalter?

## 6.2 Theoretischer und empirischer Hintergrund

In diesem Abschnitt werden (1) unterschiedliche Modelle zur Übertragung von Tests vorgestellt werden, gefolgt von (2) Studien zur Adaption von Tests für gesprochene Sprachen. In einem nächsten Schritt (3) wird sowohl das Thema der Adaption von Gebärdensprachtests als auch der Test vorgestellt, der als Vorlage zur Adaption in die DGS dient. (4) Abschließend werden Studien zum Gebärdenspracherwerb und zu den Strukturen, die in der Testvorlage abgebildet sind, ausgewertet werden.

### 6.2.1 Modelle zur Übertragung von Tests in andere Kulturen und Sprachen

Studien zu kulturübergreifenden Testadaptionen schlagen drei verschiedene Modelle einer Testübertragung von einem Ausgangs- in einen Zieltest abhängig von der Übereinstimmung beziehungsweise Überschneidung des abgebildeten Konstrukts in der Ausgangs- und Zielkultur vor (van de Vijver und Leung, 1997a, 1997b; van de Vijver und Poortinga, 2005). Die Modelle (*application*, *adaptation* und *assembly*) unterscheiden sich in dem Maß, in dem (1) Testitems einfach nur übersetzt (*application*) und (2) teilweise verändert und an die Zielkultur angepasst werden (*adaptation*) sowie ob (3) der Test komplett neu erstellt wird (*assembly*), wobei trotzdem das zugrunde liegende Konstrukt das gleiche bleibt beziehungsweise ist.

Der Ansatz der Testadaption wird zur Testübertragung vorgeschlagen, da grundsätzlich das gleiche abgebildete Konstrukt in dem Ausgangs- und Zieltest erfasst wird, das heißt *Sprachentwicklung*. Der Ansatz der Testadaption wurde bei der Auswertung der Literatur bei der Übertragung von einem Ausgangs- in einen Zielsprachtest als erfolgreicherer Ansatz – im Vergleich zu einer Übersetzung – herausgearbeitet (Friend und Keplinger, 2008; Hamilton et al., 2000; Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir und Ellis Weismer, 1996).

### 6.2.2 Adaption von Lautsprachtests

Der Begriff der *Adaption* beschreibt den ganzen Prozess der Übertragung von einem Ausgangs- bis hin zu einem Zieltest unter Berücksichtigung kultureller und linguistischer Aspekte. Der Begriff der *Übersetzung* beschreibt entweder einen Teilschritt des Adaptionsprozesses oder beschränkt sich sehr stark auf eine sehr enge Übertragung von einem Test in der Ausgangs- in die Zielsprache (Geisinger, 1994; Hambleton, 1994, 2005; Hambleton und Patsula, 1998).

Die Ergebnisse von Studien zu der *Übersetzung* eines Ausgangssprachtests in einen neuen Zielsprachtest waren weniger erfolgreich (zum Beispiel Alant und Beukes, 1986; Chavez, 1982; Rosenbluth, 1976; Simon und Joinier, 1976) als die *Adaption von Sprachtests* (Friend und Keplinger, 2008; Hamilton et al., 2000; Jackson-Maldonado et al., 1993; Maital et al., 2000; Thordardottir und Ellis Weismer, 1996). Die erwähnten Testadaptionen beziehen sich in den meisten Fällen auf das MacArthur-Bates Communication Development Inventory (Fenson et al., 1993). Dies ist eine Sprachentwicklungs-Checkliste, die von Eltern zur Erfassung des Wortschatz- und frühen Grammatikerwerbs von Kindern im Alter von 8 bis 30 Monaten ausgefüllt wird. Die Hauptmerkmale bei der Testadaption von gesprochenen Sprachen waren, dass (1) Ergebnisse aus Studien über den Erwerb und die Struktur von Sprachen einbezogen wurden und (2) Testitems, die kulturell irrelevante Konzepte der Zielkultur darstellten, entfernt wurden und neue Testitems hinzugefügt wurden, die für die Zielkultur relevant waren. (3) Des Weiteren wurde in einigen Fällen eine Expertengruppe (bestehend aus Lehrern, Forschern, Eltern) in die Testadaption einbezogen.

#### 6.2.2.1 Adaption von Gebärdensprachtests

Die Ausgangslage für die *Adaption von Gebärdensprachtests* ist eine ganz andere: Es ist oft nicht möglich, sich auf Forschungsergebnisse von Gebärdensprachen zu beziehen, wie es für die oben dargestellten Studien zu gesprochenen Sprachen möglich ist. Auch bei einem adaptierten Test ist es wichtig, die Reliabilität und die Validität zu untersuchen (Hambleton, 1994, 2001, 2005). Durch die Auswertung der Literatur über die Adaption von Gebärdensprachtests konnten zwei große Problembereiche herausgefunden werden: (1) sprachspezifische Strukturen und (2) kulturelle Themen.

Gebärdensprachtests können je nach Zielsetzung und Anwendungsbereich drei Kategorien zugeordnet werden (Haug, 2008a): (1) Tests für Forschungszwecke, (2) Tests zur Auswertung von bilingualen Schulversuchen und (3) Tests zur Erfassung der Gebärdensprachentwicklung gehörloser Kinder. Hier werden nur einige Tests der letzten Kategorie vorgestellt werden[37]. In dieser Kategorie befinden sich Tests mit dem Ziel, die Gebärdensprachentwicklung gehörloser Kinder zu erheben, um gegebenenfalls eine gebärdensprachliche Intervention einleiten zu können. Tests in dieser Kategorie testen (1) entweder Sprachverständnis oder -produktion oder beides und sie fokussieren (2) auf eine breite Altersspanne (je nach Test) von 8 Monaten bis zu 15 Jahren. Des Weiteren testen sie (3) meistens in speziellen linguistischen Bereichen, wie prä-linguistische Kommunikation, Phonologie, lexikalisches Wissen, Morphologie bis hin zu Syntax und narrativen Strukturen. Die meisten Tests schauen aber auf spezifische Strukturen, wie beispielsweise auf morphosyntaktische Strukturen. Tests in dieser Kategorie wurden für ASL, BSL, Australische Gebärdensprache (Auslan: Australian Sign Language), DGS und die Niederländische Gebärdensprache entwickelt (Anderson und Reilly, 2002; Baker und Jansma, 2005; Bizer und Karl, 2002; Fehrmann et al., 1995a, 1995b; Herman et al., 2004; Herman et al., 1999; Hermans et al., 2010; Hoiting, 2009; Jansma et al., 1997; Johnston, 2004; Maller et al., 1999; Mounty, 1993, 1994).

Neben den bereits verfügbarem Vokabeltest für die DGS (Perlesko) gibt es noch zwei weitere, allerdings nicht veröffentlichte Tests zur DGS. Der eine, der Aachener Test zu Gebärdensprache (ATG; Fehrmann et al., 1995a, 1995b), ist ein kriteriumsorientierter Test zur DGS, der recht komplex und zu lang ist, um in Schulen effizient angewendet zu werden. Der andere ist der Computertest zur DGS (CTDGS), der zu einer Studie, die den Zusammenhang zwischen DGS- und Deutschkompetenz ermitteln sollte, entwickelt wurde (Mann, 2008).

Von all diesen Tests (über alle Gebärdensprachen hinweg) sind die wenigsten veröffentlicht: Es sind zwei Tests zur BSL (Herman et al., 1999; Herman et al., 2004), ein Test zur Niederländischen Gebärdensprache (Hermans et al., 2010) und ein Vokabeltest zur DGS (Bizer und Karl, 2002) veröffentlicht worden. Diese geringe Zahl an veröffentlichten Tests deutet auch auf eine „Schwäche“ von Gebärdensprachtests ganz allgemein hin,

[37] Für einen Überblick über die verschiedenen Gebärdensprachtests siehe Haug (2008 a) oder www.signlang-assessment.info.

nämlich dass sie noch keine nachgewiesenen psychometrischen Eigenschaften haben, die sie aber zur Rechtfertigung ihrer Veröffentlichung aufweisen müssen (Haug, 2008a). Alle diese Tests sind noch in Entwicklung und daher trotz des ausgewiesenen Bedarfs nicht verfügbar. In dieser Situation spiegelt sich auch die nicht zufriedenstellende Forschungslage wider.

Aufgrund der bisher dargestellten Tests werden drei Kriterien definiert, die zur Auswahl einer Testvorlage angewendet werden, die als Grundlage für die Adaption in die DGS dient: Der Test sollte (1) gute psychometrische Eigenschaften aufweisen, (2) Sprachverständnis erfassen und (3) für Kinder ab 3 Jahren geeignet sein. Die Kriterien treffen nur auf den *BSL Receptive Skills Test* (BSL RST) zu (Herman et al., 1999).

Der BSL RST ist ein videobasierter Verständnistest für gehörlose Kinder im Alter von 3 bis 11 Jahren. Er testet morphosyntaktische Strukturen auf der Verständnisebene. Der BSL RST besteht aus zwei Teilen. Der erste Teil beinhaltet eine Vokabelüberprüfung (Vortest), die einerseits die Kenntnisse der Vokabeln (22 einfache Nomen), die später im eigentlichen Test vorkommen, ermittelt. Andererseits wird die Vokabelüberprüfung durchgeführt, um mögliche regionale Varianten, die das Kind benutzt, erkennbar zu machen (es gibt eine nord- und eine südenglische Version des BSL RST). Die Kinder sind dazu angehalten, das zu gebärden, was sie auf den Bildern sehen. Der eigentliche BSL RST besteht aus 40 Multiple-Choice Items. Die Kinder sehen eine BSL-Sequenz auf Video und können dann in einem Heft zwischen 3 bis 4 möglichen Antworten (Bildern) eine auswählen. Der BSL RST testet (1) räumliche Verbmorphologie (*agreement verbs*, *AB verb constructions* und *spatial verbs with whole entity classifiers*),[38] (2) Anzahl und Distribution, (3) Verneinung, (4) *size and shape specifiers (SASS),* (5) *handling classifiers* und (6) morphologisch abgeleitete Nomen-Verb-Paare (Herman, 2002; Herman et al., 1999). Der Test wurde an 138 gebärdensprachkompetenten gehörlosen und hörenden Kindern standardisiert und zeigt gute

---

[38] Englische Fachbegriffe, die sich nur schwer ins Deutsche übersetzen lassen, da sie immer mit einem Modell oder Theorie verbunden sind, werden nicht übersetzt werden. Es sollen aber an dieser Stelle mögliche Übersetzungsvorschläge präsentiert werden: (1) *agreement verbs*: Richtung-, Kongruenz-, Übereinstimmungs- oder Transferverben; (2) *AB verb constructions*: relativ neuer Begriff aus der Forschung zu BSL (Morgan und Woll, 2002 b), siehe Erklärung im Text; (3) *spatial verbs with whole entity classifiers*: Substitutorverben; (4) *size and shape specifiers (SASS):* Größe-Form-Klassifikatoren, Skizze+Maß; und (5) *handling classifiers:* Handhabungsklassifikatoren, Manipulatorverben.

psychometrische Eigenschaften auf. Zu Arbeitszwecken in der vorliegenden Studie werden die gleichen (zum Teil englischen) Fachbegriffe wie in der britischen Studie verwendet.

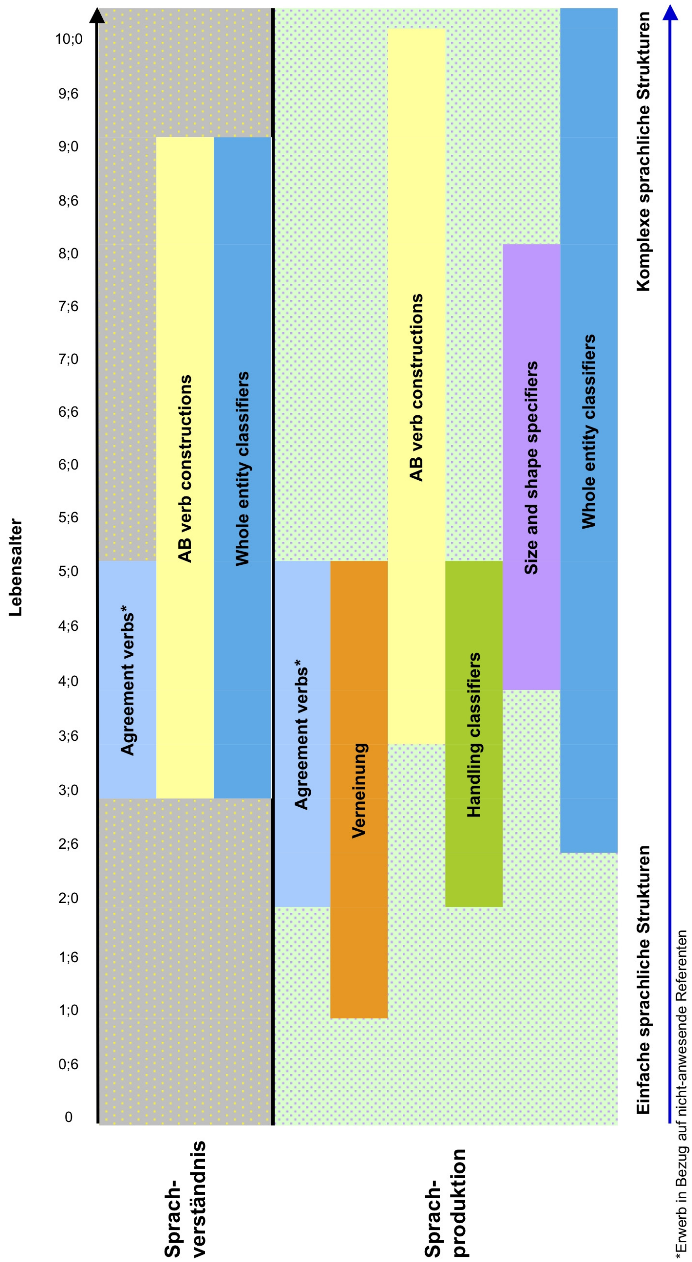
### 6.2.3 Linguistische Studien zur Testadaption

#### 6.2.3.1 Studien zum Gebärdenspracherwerb

Eine wichtige Grundlage zur Adaption eines DGS-Verständnistests sind Spracherwerbsstudien. Da es zur DGS nur eine Studie gibt (Hänel, 2003), wurden Spracherwerbsstudien aus anderen Gebärdensprachen herangezogen, um einen Überblick über den Erwerb der sprachlichen Strukturen, die im BSL RST getestet werden, zu gewährleisten. Es wurden schwerpunktmäßig Studien herangezogen, die sich auf die dem DGS-Test entsprechende Altersgruppe der 4- bis 8-Jährigen beziehen. Diese Studien beziehen sich auf die ASL (Anderson und Reilly, 1997, 2002; Bellugi et al.; 1988, Hoffmeister, 1992; Martin und Sera, 2006; Reilly, 2006; Reilly und Anderson, 2002; Schick, 1987, 1990; Slobin et al., 2003), Auslan (de Beuzeville, 2004, 2006), BSL (Morgan et al., 2003, 2006, 2008; Morgan und Woll, 2002b, 2003), NGT (Slobin et al., 2003), Brasilianische (Bernardino, 2005) und Italienische Gebärdensprache (Pizzuto, 2002).

Alle ausgewerteten Erwerbsstudien zusammen genommen ergaben eine Übersicht darüber, in welcher Abfolge welche sprachlichen Strukturen in der Entwicklung auftauchen (*emergence*) und wann die Entwicklung abgeschlossen ist (*mastery*). Die Abbildung auf der nächsten Seite stellt diesen Prozess der Entwicklung der im Test abgebildeten sprachlichen Strukturen grafisch dar (Abbildung 6.1). Die Auswertung der Studien zum Gebärdenspracherwerb wurde später – zusammen mit anderen sprachwissenschaftlichen Studien (nächster Abschnitt) – als Grundlage zur Herleitung von Hypothesen genutzt, die wiederum eine wichtige Grundlage im Adaptionsprozess waren.

Abbildung 6.1: Überblick zum Gebärdenspracherwerb basierend auf den ausgewerteten Studien



#### 6.2.3.2 Studien zu DGS-Strukturen

Die Auswertung der Literatur über den Gebärdenspracherwerb hatte das Ziel, den Entwicklungsaspekt der dargestellten sprachlichen Strukturen abzubilden. Es wurden auch Studien über die Strukturen, die im BSL-Test vorkommen, in der DGS ausgewertet, um sprachliche Unterschiede zwischen BSL und DGS herauszufinden. Die Studien zur DGS sollen hier kurz zusammengefasst werden.

(1) *Räumliche Verbmorphologie:* In der Kategorie räumliche Verbmorphologie im BSL RST befinden sich drei verschiedene sprachliche Strukturen, die hier getrennt dargestellt werden sollen (*agreement verbs, AB verb constructions, whole entity classifiers*).

(2) *Agreement verbs* werden in der DGS in ähnlicher Weise realisiert wie in anderen Gebärdensprachen, das heißt durch die Bewegung durch den Raum. Der Anfangspunkt eines *agreement verbs* bezeichnet den Ort des Subjektes und der Endpunkt das Objekt (Glück, 2001). Es gibt aber auch DGS-spezifische Strukturen, zum Beispiel den *Person agreement marker* (PAM) AUF, der in Fällen, wo beispielsweise ein *agreement verb* aus phonetischen Gründen nicht durch den Anfangs- und Endpunkt der Bewegung die Kongruenz zwischen dem Subjekt und dem Objekt ausdrücken kann, benutzt wird. Ein Beispiel hierzu ist das DGS-Verb HASSEN (Mathur und Rathmann, 2001; Papaspyrou et al., 2008; Rathmann, 2003; Rathmann und Mathur, 2002).

(3) Die Kategorie von *AB verb constructions* stammt aus Erwerbsstudien zur BSL (Morgan und Woll, 2002b), die auch im BSL RST überprüft werden. Zu Arbeitszwecken wurde dieser Begriff auch in dieser vorliegenden Studie angewendet. Ein Beispielsatz ist *Das Mädchen kämmt dem Jungen die Haare.* Die Handlung wird von zwei verschiedenen Perspektiven aus realisiert, im ersten Teil wird *Das Mädchen kämmt* und im zweiten Teil *kämmt dem Jungen die Haare* realisiert. Es sind immer ein Agens (Mädchen), ein Patiens (Junge), eine Handlung (kämmen) und ein betroffener Körperteil, an dem die Handlung ausgeführt wird (Haare), involviert (Morgan und Woll, 2002b, 2003). Die „Bestandteile" von *AB verb constructions* existieren auch in der DGS. Der linguistische Status von *AB verb constructions* in der DGS, das heißt ob sie wirklich eine eigene Kategorie von Verben darstellen, ist ungeklärt.

(4) *Whole entity classifiers* gibt es gleichermaßen in der DGS. Bei *whole entity classifiers* steht die Handform stellvertretend für eine Klasse von Nomen, auf die sich die Handlung des Referenten bezieht. Die B-Handform kann sich beispielsweise auf ein Auto beziehen, das sich entweder durch den Raum bewegt oder sich an einem Ort befindet, und stellt so die Bewegungen und Orte des Autos in der „wirklichen" Welt dar (Glück, 2001; Glück und Pfau, 1997 a, 1997 b, 1998; Happ und Vorköper, 2005; Perniss, 2007).

(5) *SASS* gibt es gleichermaßen in der DGS. Sie beschreiben häufig Form, Muster und Ausdehnung eines Objektes und drücken ähnlich den gesprochenen Sprachen adjektivische Informationen aus (Glück, 2001, 2005; Glück und Pfau 1997a, 1997b; Happ, 2005).

(6) *Handling classifiers* sind gleichermaßen vertreten in der DGS wie in anderen Gebärdensprachen, da sie darstellen, wie die menschliche Hand das jeweilige Objekt (abhängig von der Form) oder den jeweiligen Gegenstand handhabt (Glück, 2001; Glück und Pfau 1997a, 1997b, 1998; Happ und Vorköper, 2005).

(7) *Anzahl und Distribution* beschreiben unter anderem unterschiedlichste Formen der Pluralbildung in Gebärdensprachen. Diese Struktur gibt es in gleicher Weise in der DGS, nur die Art der Realisierung kann im Speziellen etwas anders sein wie in BSL (Perniss, 2001; Pfau und Steinbach, 2005, 2006). Beispielsweise wird ein unspezifischer Plural für die nominale Gebärde HAUS++ durch eine Wiederholung am gleichen Ort realisiert (Perniss, 2001). BSL hingegen realisiert diese Form der Pluralbildung durch eine Seitwärtswiederholung der nominalen Gebärde.

(8) *Verneinung* in der DGS verfügt über andere lexikalische Verneinungsmarker im Vergleich zu BSL. Teilweise wird die Verneinung wie in anderen Gebärdensprachen durch eine Kombination von manuellen und nichtmanuellen Komponenten realisiert (Pfau, 2001, 2004; Pfau und Quer, 2002, 2007).

(9) *Morphologisch abgeleitete Nomen-Verb-Paare* scheint es in DGS nicht zu geben (Becker, 2003).

Zusammenfassend lässt sich festhalten, dass sprachliche Strukturen, die es in der BSL gibt, in einer ähnlichen Art, wenn auch anders realisiert, auch in der DGS vorkommen (zum Beispiel Verneinung). Es gibt aber auch sprachliche Strukturen, die sehr sprachspezifisch sind, zum Beispiel der PAM in

DGS. Wiederum gibt es sprachliche Strukturen, die weniger sprachspezifische Unterschiede aufweisen, zum Beispiel *whole entity classifiers.* Dies ist möglicherweise verursacht durch die Art der Darstellung von Objekten und Ereignissen im Raum (Sandler und Lillo-Martin, 2006; Schembri, 2003).

Sowohl Unterschiede als auch Ähnlichkeiten zwischen Gebärdensprachen stellen wichtige Erkenntnisse dar, die bei der Testadaption einfließen werden.

### 6.2.4 Gebärdenspracherwerb und Testadaption

Aufgrund der vorher ausgewerteten Studien über den Erwerb von Gebärdensprachen, über die DGS als auch über weitere sprachvergleichende Studien wird hier argumentiert, dass in Anbetracht der Forschungslage zur DGS (1) Spracherwerbsstudien herangezogen werden können, um den Entwicklungsaspekt in dem adaptierten DGS-Test abzubilden. Zudem werden auch (2) Studien über DGS-spezifische Strukturen und (3) sprachvergleichende Studien herangezogen, die darüber Aufschluss geben, welche Strukturen in dem adaptierten DGS-Test abgebildet sein sollen.

Auf der Grundlage dieser drei schwerpunktmäßig ausgewerteten Studien werden folgende Hypothesen formuliert:

(1) *Sprachübergreifende Erwerbsstudien legen nahe, dass sowohl einfache als auch komplexere Strukturen in Bezug zum Alter stehen und so Entwicklung widerspiegeln.*

(2) *Diese einfachen und komplexeren Strukturen sind in einfacheren und komplexeren Testitems abgebildet, die wiederum einen Hinweis auf den Spracherwerb geben sollen.*

(3) *Aufgrund der Studien zu DGS-Strukturen und sprachübergreifenden Strukturen werden sprachspezifische Strukturen in diesen Testitems abgebildet.*

Diese Hypothesen dienen als Grundlage zur Operationalisierung und geben die Richtlinie dafür an, welche Testitems in der adaptierten DGS-Testversion abgebildet werden. Eine „visuelle Darstellung" dieser Hypothesen ist in der Abbildung *Map Ranking of Item Complexity* (Figure 2.2) dargestellt. Diese Hypothesen fließen als Grundlage in den methodischen Teil der Testadaption ein und werden zumeist implizit, teilweise aber auch explizit überprüft.

## 6.3 Methodik

In diesem Abschnitt werden (1) das Studiendesign, (2) die Beschreibung der Versuchspersonen, (3) die Datenerhebung und die (4) Auswertungsmethoden beschrieben. Das Forschungsdesign in dieser Studie folgt einem experimentellen Ansatz in Form einer Querschnittsstudie, in deren Rahmen die Testadaption von einer Gebärdensprache in eine andere durchgeführt wird.

### 6.3.1 Studiendesign

Grundlage für die Entwicklung des Erhebungsinstruments ist der BSL RST (Herman et al., 1999), ein standardisierter Sprachverständnistest zur Britischen Gebärdensprache, der in die DGS adaptiert wurde. Dieser Prozess der Adaption besteht aus mehreren Schritten, die in der Tabelle 6.1 dargestellt werden.

Der adaptierte und im Gegensatz zum BSL-Test computerbasierte DGS-Verständnistest (DGS-VT) bestand aus 21 Vokabeln für die Vokabelüberprüfung (eine Vokabel, das *Hörgerät*, wurde in der adaptierten Version entfernt) und 53 Testitems (3 Übungsitems, 40 adaptierte Testitems, 10 neu entwickelte Testitems). Die 40 adaptierten Testitems folgten der im BSL-Test verwendeten Reihenfolge, wobei die Testitems nach ihrem Schwierigkeitsgrad angeordnet waren. Die 10 neu entwickelten Testitems folgten nach. Durch das standardisierte Testformat konnten die Testergebnisse automatisch auf dem Laptop abgespeichert werden, während beim BSL-Test ein Tester die Antworten der Kinder auf einem Auswertungsbogen ankreuzte. Dadurch entsteht auch ein im hohen Maße einheitliches Format der Testdurchführung, das für alle Kinder die gleichen Testbedingungen mit sich bringt. Das Format der Testitems war Multiple-Choice (Mehrfachauswahl), das heißt die Kinder konnten aus 3 bis 4 Antwortmöglichkeiten die richtige Antwort auswählen. Neben der richtigen Antwort wurden 2 bis 3 alternative Antworten angeboten, so genannte Distraktoren oder „Ablenker". Die Abbildungsformen der Antwortmöglichkeiten waren leicht verständliche, auf Kinder abgestimmte farbige Zeichnungen, die auf das Wesentliche reduziert waren.

Tabelle 6.1: Überblick über den Prozess der Testadaption

| *Schritte* | *Beschreibung der Schritte* |
|---|---|
| 1. Durchsicht und Überarbeitung der Testmaterialien | Die Testmaterialien (Bilder) wurden durchgesehen und überarbeitet, der rote und runde britische Briefkasten auf einem Bild wurde zum Beispiel in der Adaption durch den deutschen eckigen und gelben Briefkasten ersetzt. |
| 2. Voruntersuchung 1 | Um die Angemessenheit der Testitems zu überprüfen, wurden in drei Regionen Daten (Dialektvarianten) erhoben. Es wurden bei 5 der 22 Vokabeln der Vokabelüberprüfung Dialektvarianten festgestellt, die aber nicht eindeutig einer Region zugeordnet werden konnten. |
| 3. Übertragung der Testitems | (1) Festlegung der Reihenfolge der Testitems (wie im BSL-Test)<br>(2) Auswertung der Literatur, um herauszufinden, ob es die zu testenden sprachlichen Strukturen in der DGS auch gibt<br>(3) Entwicklung von 10 zusätzlichen Testitems |
| 4. Aufnahme der Testitems | Aufnahme der Testeinführung und -items mit einem Gehörlosen |
| 5. Programmierung der Testoberfläche | Programmierung einer benutzerfreundlichen Testoberfläche, die auf einem Laptop läuft und wo auch die Ergebnisse automatisch gespeichert werden können |
| 6. Voruntersuchung 2 | Die erste Testversion wurde mit 2 Gruppen getestet:<br>(1) hörende Kinder, die nicht Gebärdensprache konnten;<br>(2) gehörlose Erwachsene. |
| 7. Überarbeitung der ersten Testversion | Überarbeitung der ersten Testversion basierend auf den Ergebnissen der Voruntersuchung 2:<br>(1) Änderungen an den Bildern;<br>(2) nochmalige Aufnahme von einigen Items;<br>(3) Veränderungen der Testoberfläche. |

Zu den lexikalischen Varianten: Um nicht mehrere Testversionen entwickeln zu müssen, wurde entschieden, vor dem Beginn des Verständnistests einen „Übungsteil" einzubauen. Bei Letzterem zeigte das Sprachmodell nach der Einblendung eines Bildes die entsprechende Gebärde, wie sie später im Test benutzt wurde. Dies betraf fünf lexikalische Gebärden (JUNGE, KIND, HUND, MUTTER, TEDDYBÄR). Ziel dieses Übungsteils war es, das Kind erkennen zu lassen, welche Gebärden nachher im Verständnistest benutzt werden. Die Testanleitung so wie alle anderen Teile des Tests lagen in einem einheitlichen Format auf Video vor.

Der Test bestand aus drei Teilen. Bevor der Test anfing, konnte der Tester eine dem Kind zugewiesene Identifikationsnummer (ID) eingeben. Der erste Testteil bestand aus einer allgemeinen Einleitung über den Test und seinen Aufbau, gefolgt von der Vokabelüberprüfung, in der die Kinder die Bilder von Nomen sahen und aufgefordert waren zu gebärden, was sie sahen. Die Ergebnisse wurden auf einem gesonderten Auswertungsbogen eingetragen (Appendix E-1) und auch auf Video aufgenommen. Wenn das Kind eine Gebärde nicht wusste, ging der Tester zu dieser Vokabel zurück, um sicherzustellen, dass das Kind die Gebärde auch kennt. Der zweite Teil beinhaltete die oben erwähnte Übungssequenz und der dritte Testteil den eigentlichen Verständnistest. Das zweiseitige Layout wies links sowohl das Videobild, in dem das Sprachmodell die Testitems gebärdete, als auch die Knöpfe zum Navigieren auf (Stopp, Zurück, Abspielen) und rechts die drei bis vier Antwortmöglichkeiten (Figure 3.6). Das Video wurde nicht automatisch gestartet, sondern musste vom Kind, durch Anklicken des Bildes gestartet werden. Nach dem Anschauen des Videos konnte das Kind eine Antwort auswählen; dabei erschien ein Pfeil, der zum nächsten Testitem führte. Das Kind hatte auch die Möglichkeit, sich für ein anderes Bild zu entscheiden. Sobald allerdings der grüne Pfeil/Knopf gedrückt wurde, war das zuletzt angeklickte Bild als Ergebnis gespeichert. Es war nicht möglich, zu einem Testitem zurückzukehren. Die Videos konnten maximal zweimal angeschaut werden. Es besteht aber die Möglichkeit, diese Voreinstellung zu ändern.

#### 6.3.1.1 Voruntersuchung: Überarbeitung des Tests

Nach der Voruntersuchung mit gehörlosen Erwachsenen wurde der Test nochmals überarbeitet. Die Änderungen betrafen insgesamt 11 Testitems. Sie bezogen sich auf (1) die Überarbeitung von Bildern, (2) die Neuaufnahme von Testitems und (3) leichte Änderungen an der Benutzeroberfläche, die zu einer einfacheren Navigation führten (eine vollständige Liste an Änderungen siehe Anhang Appendix F-3). Eines der Ergebnisse der Voruntersuchung mit gehörlosen Erwachsenen war auch, die Testitems zu den morphologisch abgeleiteten Nomen-Verb-Paaren zu entfernen. Die Voruntersuchung mit hörenden Kindern führte zu Veränderungen der Benutzeroberfläche, die zielgruppengerechter gestaltet wurde (Figure 3.7)[39].

#### 6.3.1.2 Fragebögen

Neben dem adaptierten Test wurden drei verschiedene Fragebögen benutzt. Zwei davon wurden von den Lehrpersonen und einer von den Eltern der Kinder ausgefüllt. Alle mit den Fragebögen und durch den Test erhobenen Informationen, wurden anonymisiert.

*Schülerfragebogen* (Appendix G-2): Dieser Fragebogen wurde von den Lehrpersonen ausgefüllt. In diesem Schülerfragebogen ging es um Hintergrundinformationen über den Schüler, Geburtsdatum, Lebensalter bei der Diagnose der Hörschädigung, Grad der Hörschädigung, Zeitpunkt des Beginns des Gebärdenspracherwerbs, Hörstatus und benutzte Sprachen der Eltern, Kontakt zu Gehörlosen außerhalb der Schule und um eine Einschätzung der DGS-Kompetenz des Kindes (Verständnis und Produktion).

*Lehrerfragebogen* (Appendix G-3): In diesem Fragebogen ging es um die Hintergrundinformationen über die Lehrpersonen, wie Ausbildung, Hörstatus und Wahl der Kommunikationsmittel je nach Situation in der Schule. Des Weiteren ging es auch um eine Selbsteinschätzung der eigenen DGS-Kompetenz (Verständnis und Produktion).

*Elternfragebogen* (Appendix G-1): In diesem Fragebogen, der von den Eltern ausgefüllt wurde, ging es inhaltlich um ähnliche Fragen wie bei den Schülerfragebögen. Grund dafür, diese Informationen doppelt abzufragen, war ein möglichst hohes Maß an Validität zu erreichen.

### 6.3.2 Stichprobe

Die Probanden kamen aus einer von fünf Institutionen[40] für Hörgeschädigte in Deutschland, in denen (1) entweder in der kompletten Institution

[39] Eine Auswahl an Testitems des DGS-VT kann im Internet eingesehen werden: unter http://www.signlang-assessment.info/index.php/german-sign-language-receptive-skills-test.html.

(1 von 5 Institutionen) oder (2) in einer oder mehreren Modellklassen bilingual gearbeitet wurde (2 von 5 Institutionen), oder aus (3) Institutionen, in denen „manuelle Kommunikationsmittel" eingesetzt wurden (2 von 5 Institutionen). Dies beinhaltete alle möglichen Kommunikationsformen von DGS bis LBG. Durch die begrenzt zur Verfügung stehenden Ressourcen konnte diese Studie in keinem größeren Rahmen durchgeführt werden. Die Institutionen befanden sich im Süden, Süd-Westen, Westen, Norden und Osten Deutschlands. Das Ziel war anfangs, für diese Phase der Testadaption eine möglichst homogene Gruppe von DGS-kompetenten Kindern zu testen, das heißt die Schulen wurden gebeten, möglichst nur gehörlose Kinder gehörloser Eltern in diese Studie einzubeziehen. Dieser Ansatz funktionierte nur bedingt, da einige Institutionen wollten, dass alle Kinder in der angegebenen Altergruppe getestet werden.

Zwischen Februar und Juni 2006 wurden insgesamt 74 Kinder aus diesen fünf Institutionen getestet. Das Alter der Kinder reichte von 3;9 bis 10;10 Jahre (*M* = 7;0). Die Ergebnisse von 20 Kindern wurden nicht ausgewertet, da (1) 14 Kinder nicht den kompletten Test gemacht hatten und (2) bei 6 Kindern eine andere Behinderung neben der Hörschädigung vorlag. Die verbleibende Gruppe bestand aus 54 Kindern, 29 davon männlich und 25 weiblich (Table 4.1). Von diesen 54 Kindern kamen 34 (63 %) aus gehörlosen Familien, 20 (37 %) Kinder kamen aus hörenden Familien.

Zur Hörschädigung lagen folgende Informationen vor: (1) 1 Kind mit einer leichtgradigen Schwerhörigkeit (25–40 dB), (2) 2 Kinder mit einer mittelgradigen Schwerhörigkeit (40–70 dB), (3) 29 Kinder mit einer hochgradigen Schwerhörigkeit (70–100 dB) und (4) 18 gehörlose Kinder (> 100 dB). Für 4 Kinder lag keine Information vor.

*Gehörlose Kinder aus gehörlosen Familien:* Diese Untergruppe bestand aus 34 Kindern (19 männlich, 15 weiblich) im Alter von 3;9–10;10 Jahren (*M* = 6;10) (Table 4.3).

[40] An fast allen Institutionen war die Alterspanne 4–8 (Kindergarten, Grundschule) vertreten. Nur an einem Standort waren Schule und Kindergarten zwei separate Institutionen, die zu Zwecken der Darstellung der Stichprobe zusammengenommen wurden. Diese unterschieden sich nicht in den Angaben zu ihrer Benutzung von DGS. Da keine Auswertung von einzelnen Institutionen vorgenommen wurde, stellt das Zusammennehmen beider Institutionen kein Problem dar.

*Gehörlose Kinder aus hörenden Familien:* Diese Untergruppe bestand aus 20 Kindern (10 männlich, 10 weiblich) im Alter von 5;2–9;6 Jahren (*M* = 7;4) (Table 4.4).

### 6.3.3 Datenerhebung

Mehrere Institutionen in Deutschland wurden in einem ersten Schritt mit einer Beschreibung des Projektes schriftlich angefragt, ob sie Interesse hätten, an dieser Studie teilzunehmen. Bei einem grundsätzlichen Interesse von Seiten der Schule wurde das Projekt durch den Autor persönlich in den Schulen vorgestellt. Wenn weiterhin Interesse an einer Teilnahme bestand, wurde in einem nächsten Schritt – je nach Bundesland – ein Antrag mit allen nötigen Unterlagen bei der zuständigen Behörde (zum Beispiel Kultusministerium) eingereicht. Erst wenn ein positiver Bescheid durch die Behörde vorlag, konnte mit der Rekrutierung der Probanden begonnen werden.

In einem nächsten Schritt wurden Unterlagen mit Informationen über die Studie für die Eltern zusammengestellt. Sie bestanden aus einer Projektbeschreibung, dem Elternfragebogen und einer Einverständniserklärung und wurden durch die Lehrer an die Kinder verteilt. Es war eigentlich vorgesehen, nur Kinder im Alter von 4 bis 8 Jahren in diese Studie einzubeziehen. Die Altersspanne wurde aber wegen der geringen Anzahl an Kindern etwas ausgeweitet. Nur Kinder, für die eine unterschriebene Einverständniserklärung vorlag, wurden im Rahmen der Studie getestet.

Das Testen der Kinder fand zwischen Februar und Juni 2006 statt. Die Kinder wurden einzeln vormittags aus dem Unterricht geholt zum Testen. Für Kinder im Alter von 3;9 bis 5;6 war die Testdauer circa 30 Minuten, für Kinder > 5;7 circa 20 Minuten.

Das Testen fand in allen Institutionen in einem ruhigen und abgelegenen Raum statt. Für das Testen standen ein Tisch, zwei Stühle, der Laptop, eine Computer-Maus, eine Kamera und ein Ausdruck der Testitems zur Verfügung. Die Kinder konnten selber wählen, ob sie eine externe Maus benutzen wollten (nicht alle Kinder konnten mit einer Maus umgehen). Wenn sie nicht mit einer Maus umgehen konnten, führte der Autor die Maus für sie und sie konnten entweder auf den Computerbildschirm deuten oder sich auf die vorhandenen ausgedruckten Unterlagen der Testitems beziehen.

Neben dem Auswertungsbogen für die Vokabelüberprüfung wurde noch ein Beobachtungsbogen benutzt, auf dem neben der ID die Uhrzeit sowie das Alter der Kinder notiert und darüber hinaus festgehalten wurde, ob die Kinder die Maus selber benutzt haben oder nicht. Zudem gab es auf dem Beobachtungsbogen Raum für sonstige Beobachtungen.

### 6.3.4 Auswertungsmethoden

Die Forschungsfragen in Bezug auf die psychometrischen Eigenschaften des adaptierten DGS-Tests wurden nur mit der Untergruppe gehörloser Kinder gehörloser Eltern ($N$ = 34) durchgeführt. Auch wenn der adaptierte DGS-Test mit allen gehörlosen Kindern später genutzt werden sollte, war es in dieser Phase der Testentwicklung wichtig, eine möglichst homogene Gruppe mit frühem Zugang zu DGS als Modell zu haben, um später den Spracherwerb gehörloser Kinder mit späterem Zugang zur DGS damit vergleichen zu können.

Die Forschungsfragen, bei denen bestimmte Variablen (zum Beispiel Lebensalter, Hörstatus der Eltern) in Bezug zu den Rohwerten des Tests gesetzt werden, wurden mit der gesamten Stichprobe durchgeführt, um zu sehen, ob diese Variablen (zum Beispiel Hörstatus der Eltern) einen Einfluss auf die Rohwerte der Kinder haben.

Die Hintergrundinformationen über die Kinder wurden ausschließlich den durch die Lehrer ausgefüllten Schülerfragebogen entnommen. Es war teilweise offensichtlich, dass manche Eltern mit Migrationshintergrund die Fragen nicht verstanden hatten und dass aus diesem Grund in einer Schule die Lehrer den Elternfragebogen sogar selber ausfüllten.

*Statistische Grundannahmen*: Die Überprüfung der Normalverteilungen der Variablen *Rohwerte* und *Alter* ergaben, dass diese nicht normal verteilt sind (Appendix H-1 bis H-4). Daher wurden nicht-parametrische statistische Testverfahren angewendet (Kiess, 1996). Für alle statistischen Verfahren wurde das Signifikanzniveau auf $\alpha$ = .05 festgelegt (zweiseitiger Test). Darüber hinaus wurde die Effektstärke von Korrelationskoeffizienten festgelegt: (1) .10 als klein, (2) .30 als mittel und (3) .50 als groß (Bortz, 1999; Cohen, 1992).

Da keine Daten von dem BSL RST vorlagen, war es nicht möglich, einen statistischen Vergleich des BSL RST und des adaptierten DGS-Tests durchzuführen. Alle statistischen Auswertungen wurden mit SPSS durchgeführt.

## 6.4 Ergebnisse der Evaluation

Ziel dieser Studie war die Adaption des BSL RST (Herman et al., 1999) unter Berücksichtigung kultureller, linguistischer, methodischer und theoretischer Aspekte. In diesem *Ergebniskapitel* werden nur die empirischen Forschungsfragen behandelt. Die theoretisch geleiteten Forschungsfragen werden im *Diskussionskapitel* dargestellt.

### 6.4.1 Psychometrische Eigenschaften des Tests

*Itemanalyse*: Zur Ermittlung der Itemanalyse bedarf es der Berechnung des Schwierigkeitsgrades der Testitems und deren Trennschärfe (Lienert und Raatz, 1998). Zur Ermittlung des Schwierigkeitsgrades $p_i$ und der Trennschärfe $r_{it}$ der Testitems wurden Kriterien als Entscheidungsgrundlage zur Entfernung (oder Überarbeitung) von Testitems definiert. Die Items müssen folgende Kriterien erfüllen, um im Itempool zu bleiben (Fisseni, 2004; Lienert und Raatz, 1998): (1) Item mit einem Schwierigkeitsgrades $p_i$ von .25 bis .90 und (2) Item mit einer Trennschärfe-Koeffizienten $r_{it} > .25$.

Die Ergebnisse der Itemanalyse ergaben, dass 10 Items entweder aus dem Itempool entfernt oder überarbeitet werden sollten (gesamter Überblick der Itemanalyse Appendix I-1).

*Neu entwickelte Items*: Von den zehn neu entwickelten Items wurden nur neun in der Hauptstudie benutzt (das eine Item wurde nach der Voruntersuchung 2 entfernt). Von den übrigen neun Items sollten vier aufgrund der Itemanalyse aus dem Itempool entfernt oder überarbeitet werden. Die restlichen fünf Items blieben aufgrund der Ergebnisse der Itemanalyse in dem Itempool.

*Distraktorenanalyse*: Die nicht richtigen Antworten der Kinder sollten möglichst gleich unter den Alternativantworten in einem Multiple-Choice-Test verteilt sein. Es gibt zwei Hauptgründe, weswegen Distraktoren überprüft werden müssen: (1) wenn sie von den Kindern überhaupt nicht gewählt werden oder (2) wenn sie häufiger als die richtige Antwort gewählt werden (Lienert und Raatz, 1998). Für alle Distraktoren wurde sowohl der Schwierigkeitsgrad als auch die Trennschärfe ermittelt (Lienert und Raatz, 1998). Wenn folgende Kriterien nicht erfüllt sind, werden die Items entfernt oder überarbeitet werden: (1) möglichst gleiche Verteilung der Auswahl von

Distraktoren eines Items und (2) eine negative Korrelation des Trennschärfekoeffizienten bei jedem Distraktor.

Die Mehrheit der Distraktoren zeigte gute Ergebnisse. Aufgrund der Ergebnisse der Distraktorenanalyse (Appendix I-2) konnten drei Kategorien von Distraktoren, die entweder überarbeitet und entfernt werden sollten, identifiziert werden: (1) Ein Distraktor wurde häufiger ausgewählt als die richtige Antwort und zeigte eine positive Korrelation der Trennschärfe; (2) der Distraktor eines Testitems wurde überhaupt nicht ausgewählt; (3) es gab Distraktoren, die keine negative Korrelation der Trennschärfe auswiesen (und nicht häufiger als die Zielantwort ausgewählt wurden).

Die Ergebnisse der Distraktorenanalyse trugen dazu bei, dass ersichtlich wurde, welche Distraktoren (und Testitems) überarbeitet oder ganz aus dem Itempool entfernt werden sollten.

*Homogenitätsindex*: Das Ziel der Untersuchung des Homogenitätindexes ist es zu überprüfen, ob der Test das gleiche Konstrukt hat beziehungsweise durch die Testitems verschiedene Merkmalsausprägungen des Konstrukts erfasst werden (Fisseni, 2004). Zur Ermittlung des Homogenitätsindexes $H$ wurde eine Inter-Item-Korrelation gerechnet (Bortz und Döring, 2005; Fisseni, 2004). Ein Wert von .20 bis .40 wird als ein akzeptabler Wert der Homogenität eines Tests angesehen (Briggs und Cheek, 1986). Das Ergebnis zeigte einen Wert von $H = .35$ an (Bandbreite: .20–.48, Appendix I-3).

*Reliabilität*: Ein häufig angewendetes Verfahren zur Überprüfung der inneren Konsistenz eines Tests ist das Cronbachs Alpha (Lienert und Raatz, 1998). Das Cronbachs Alpha für alle 49 Testitems betrug $\alpha = .937$. Gerechnet mit den gestrichenen zehn Testitems (s. Itemanalyse) war der Wert sogar noch etwas höher mit $\alpha = .955$ und zeigte damit einen sehr guten Reliabilitätswert des Testes an. Ein Wert bis zu .70 kann als ein „annehmbarer" Wert für den Cronbachs Alpha angesehen werden (Nunnally, 1978).

### 6.4.2 Externe Einschätzung der DGS-Kompetenz

Die Einschätzung der DGS-Kompetenz der Kinder durch die Lehrer (auf einer Skala von 1 bis 6, entsprechend dem deutschen Notensystem, Appendix G-2) wurde mit deren Rohwerten verglichen. Nur für 31 der 34 Kinder gehörloser Eltern lag diese Einschätzung vor. Der Spearman'sche Rangkorrelationskoeffizient wurde benutzt. Die Korrelation zwischen der Einschätzung der rezeptiven DGS-Kompetenz und der Rohwerte zeigte einen (fast) starken Effekt ($r_s$ = .480, $p$ = .006), die Ergebnisse der expressiven DGS-Kompetenz waren etwas schwächer und zeigten einen mittleren Effekt ($r_s$ = .374, $p$ = .038).

Die Lehrer gaben auch eine Einschätzung ihrer eigenen DGS-Kompetenz ab (Appendix G-3). Nur von 36 Lehrern lagen diese Informationen vor, wovon 32 hörend und 4 gehörlos waren. Problematisch waren vor allem die niedrigen Werte der Selbsteinschätzung (1 = Minimum und 5 = Maximum) der hörenden Lehrer (Verständnis: 3.09, Produktion: 3.39) und die hohe Bandbreite der Selbsteinschätzung von 1–5. Daher sollten die oben dargestellten Ergebnisse mit Vorsicht behandelt werden.

### 6.4.3 Inhaltliche Validierung des Tests

Eine inhaltliche Validierung wird im *Diskussionsteil* dargestellt. Zum Zeitpunkt der Testung gab es noch keinen anderen standardisierten DGS-Test, um die Übereinstimmungsvalidität ermitteln zu können.

### 6.4.4 Rohwerte in Beziehung zu anderen Variablen

Für die folgenden statistischen Auswertungen wurden sowohl Kinder gehörloser als auch hörender Eltern einbezogen. Fishers Exakter Test wurde angewendet, um zu sehen, ob es eine signifikante Beziehung zwischen den Rohwerten und (1) Geschlecht des Kindes, (2) Alter des Kindes beim Zugang zu einer Gebärdensprache, (3) Hörstatus der Eltern und (4) Lebensalter des Kindes gibt. Alle Variablen außer der Variablen Geschlecht wiesen eine signifikante Beziehung zu den Rohwerten auf (Table 4.7).

*Alter des Kindes beim Zugang zu einer Gebärdensprache*: Es kann angenommen werden, dass hörende Kinder von Geburt an Zugang zu einer Sprache haben. Bei gehörlosen Kindern haben nur 5 % gehörlose Eltern (Mitchell und Karchmer, 2004). Sie haben höchstwahrscheinlich dann auch Zugang zu einer Erstsprache von früh an. Inwiefern der frühe oder späte Zugang zu einer Gebärdensprache einen Einfluss auf die Rohwerte hat, ist ein wichtiger Punkt für die Testadaption, deren Ziel es ist, eine möglichst homogene Gruppe für die Standardisierung zu definieren.

Informationen über den frühen (0–3 Jahre) und späten (3–6 Jahre) Zugang zu DGS waren nur von 35 Kindern verfügbar (Table 4.8). Die Untergruppe mit frühem Zugang umfasste 27 Kinder (21 mit gehörlosen, 6 mit hörenden Eltern) mit einem Altersdurchschnitt von 7;5 Jahren (5;3–10;10) und die Untergruppe mit spätem Zugang bestand aus 8 Kindern (alle mit hörenden Eltern) mit einem Altersdurchschnitt von 6;5 Jahren (5;2–8;1).

Eine ANOVA wurde gerechnet zum Vergleich der Rohwerte der beiden Untergruppen[41]. Die Untergruppen mit dem frühen Zugang zur DGS mit einem durchschnittlichen Rohwert von 36.04 schnitt statistisch signifikant besser ab als die Untergruppen mit dem späten Zugang mit einem durchschnittlichen Rohwert von 19.63 ($F = 28.95$, $df = 1$, $p < .001$). In einem nächsten Schritt wurde die Kontrollvariable (Kovariate) *Lebensalter* in die Auswertung einbezogen. Das Ergebnis zeigt, dass die Kovariate einen Einfluss auf die Rohwerte der beiden Untergruppen hat ($F = 8.4$, $df = 1$, $p = .007$), aber die Variable *früher Zugang zu einer Gebärdensprache* ist stärker ($F = 23.42$, $df = 1$, $p < .001$). Der Altersunterschied der Untergruppen mit frühem Zugang ($M$ = 7;5) im Vergleich zu der Untergruppen mit spätem Zugang ($M$ = 6;5) war nicht statistisch signifikant unterschiedlich ($F = 3.11$, $df = 1$, $p = .087$).

Es wurde auch versucht, die Variable *Lebensalter* zu adjustieren, das heißt die Variable *Gebärdensprachalter*[42] zu ermitteln. Leider lagen diese Informationen nur für 35 Kinder vor. Das Gebärdensprachalter konnte nicht in die Auswertung einbezogen werden, da nicht für genug Kinder Informationen darüber angegeben waren. Bei dem Versuch, beide Untergruppen für Gebärdensprachalter anzupassen, gab es nur noch eine Stichprobe von $N = 25$, die sich auf zwei sehr ungleiche Untergruppen verteilte (früher Zugang: $n = 24$; später Zugang: $n = 1$; s. Table 4.9).

---

[41] Zuerst wurde ein Mann-Whitney-U-Test angewendet. Die Ergebnisse belegen, dass der mittlere Rang der Testrohwerte der Kinder in der Untergruppe mit frühem Zugang zu einer Gebärdensprache ($n = 27$, mittlerer Rang = 21.48) statistisch signifikant höher ist als in der Untergruppe mit spätem Zugang ($n = 8$, mittlerer Rang = 6.25, $U = 14$, $p < .001$). In einem nächsten Schritt wurde eine ANOVA angewendet, die die vorherigen Ergebnisse bestätigte ($M$ früher Zugang = 36.04, $M$ später Zugang = 19.62, $F = 28.95$, $df = 1$, $p < .001$). Es wurde deswegen eine ANCOVA angewendet, um auch Kontrollvariablen (zum Beispiel Lebensalter, Gebärdensprachalter) mit einbeziehen zu können. Es gibt keine nicht-parametrischen Testverfahren unter SPSS, um eine Kontrollvariable einbeziehen zu können.

[42] *Gebärdensprachalter* beschreibt die Länge der Benutzung einer Gebärdensprache: Lebensalter abzüglich Länge der Benutzung einer Gebärdensprache. Ein 6-jähriges gehörloses Kind hörender Eltern, dass mit 3 Jahren Zugang zu einer Gebärdensprache erhielt, hat zum Beispiel ein Gebärdensprachalter von 3 Jahren. Dies wurde auch in einer Untersuchung zum Erwerb der Niederländischen Gebärdensprache zwischen gehörlosen Kindern gehörloser und hörender Eltern angewendet (Hoiting, 2009).

Zusammenfassend lässt sich festhalten, dass die Variable des frühen Zugangs ein statistisch besseres Ergebnis erzielte, aber teilweise sind die beiden Untergruppen früher und später Zugang zu einer Gebärdensprache nicht unabhängig vom Hörstatus der Eltern. Des Weiteren konnte das Gebärdensprachalter nicht ermittelt werden, um eine weitere erklärende Variable in diese Berechnung einzubeziehen.

*Hörstatus der Eltern*: Es wurde in der Literatur berichtet, dass gehörlose Kinder gehörloser Eltern bessere Gebärdensprachkenntnisse haben als gehörlose Kinder hörender Eltern (Strong und Prinz, 1997, 2000). Der Hörstatus der Eltern ist aber nicht allein erklärend für bessere Rohwerte, sondern auch der frühe Zugang zu einer Gebärdensprache ist eine wichtige Variable. Wie auch in dem obigen Abschnitt erwähnt, waren nicht ausreichend Daten über das Gebärdensprachalter vorhanden, um es in die Berechnung einzubeziehen.

Ein Mann-Whitney-U-Test wurde angewendet, um die Unterschiede in den Rohwerten der gehörlosen Kinder gehörloser Eltern ($n$ = 34) und den gehörlosen Kindern hörender Eltern ($n$ = 20) zu vergleichen. Der mittlere Rang der Testrohwerte der gehörlosen Kinder gehörloser Eltern mit 31.71 (Alter = 3;9–10;10, $M$ = 6;10) war statistisch signifikant höher als der mittlere Rang der Testrohwerte von 20.35 der gehörlosen Kinder hörender Eltern (Alter = 5;2–9;6, $M$ = 7;4, $U$ = 197, $p$ = .010). Kein statistisch signifikanter Unterschied wurde zwischen den beiden Durchschnittsaltern (6;10 vs. 7;4) der beiden Untergruppen gefunden ($U$ = 268, $p$ = .197).

Zusammenfassend lässt sich festhalten, dass eine statistisch signifikante Beziehung zwischen Hörstatus und Rohwerten hergestellt werden konnte, allerdings erklärt das nur eine Beziehung und keine Kausalität, die leider aufgrund der Daten über das Gebärdensprachalter nicht untersucht werden konnte.

*Lebensalter der Kinder*: Wichtig für die Entwicklung und Adaption von Sprachentwicklungstests ist es, das sie in Bezug auf das Lebensalter differenzieren können. Es wurde ein Rangkorrelationskoeffizient $r_s$ nach Spearman unter Einbezug der gesamten Stichprobe ($N$ = 54) zwischen der Variabel *Lebensalter* und *Rohwerten* gerechnet. Das Ergebnis zeigt einen starken Effekt ($r_s$ = .530, $p$ < .001), das heißt je älter die Kinder, desto höher die Rohwerte. Getrennt für beide Untergruppen zeigte das Ergebnis der Korrelation zwischen gehörlosen Kinder gehörloser Eltern einen noch stärkeren Effekt ($r_s$ = .81, $p$ < .001); der Effekt für die gehörlosen Kinder hörender Elter war etwas niedriger als die der gehörlosen Eltern ($r_s$ = .541, $p$ = .014), aber dennoch zeigte es einen starken Effekt.

### 6.4.5 Zusammenfassung der Ergebnisse

Zusammenfassend lässt sich festhalten, dass der adaptierte DGS-Test gute psychometrische Eigenschaften aufweist. Die Ergebnisse der externen Variable (Einschätzung der DGS-Kompetenz der Kinder durch die Lehrer) müssen mit Vorsicht behandelt werden, da die selbst geschätzte DGS-Kompetenz der hörenden Lehrer recht unterschiedlich ist.

Die Variablen Alter des Kindes beim Zugang zu einer Gebärdensprache, Hörstatus der Eltern und Lebensalter des Kindes gaben Informationen, um unterschiedliche Aspekte, die die Testrohwerte der Kinder beeinflussen, zu erklären. Dies sind wichtige Informationen zur Weiterentwicklung des adaptierten DGS-Tests.

## 6.5 Diskussion

Das Ziel dieser Studie war es, kulturelle, linguistische, methodische und theoretische Themen der Testadaption von BSL in DGS zu untersuchen.

### 6.5.1 Erkenntnisse aus der Evaluation des adaptierten Tests

#### 6.5.1.1 Kulturelle Aspekte der Testadaption

Kulturelle Aspekte waren weniger stark vertreten als beispielsweise linguistische oder methodische Themen in der Testadaption. Dies bezog sich vor allem auf die Darstellung kultureller Konzepte in den Testmaterialien. So wurde zum Beispiel der rote und runde britische Briefkasten durch den gelben und eckigen Briefkasten ersetzt. Dieses Problem trat auch in der Adaption anderer Gebärdensprachen auf (zum Beispiel Dänische Gebärdensprache; Haug und Mann, 2008).

#### 6.5.1.2 Psychometrische Eigenschaften des adaptierten DGS-Tests

Die Ergebnisse der *Itemanalyse* gaben Aufschluss über die Entfernung oder Überarbeitung von Testitems (Appendix J-1). Es entstand so auch eine neue Rangordnung der Items im Vergleich zum BSL-Test. Die *neu entwickelten Testitems* zeigten teilweise gute Ergebnisse und werden deshalb so in dem Itempool bleiben. Die *Distraktorenanalyse* gab Aufschluss darüber, welche Distraktoren bestimmter Items überarbeitet werden mussten beziehungsweise zusätzliche Hinweise darauf, bestimmte Items nicht zu entfernen, sondern in überarbeiteter Form im Test zu belassen. Sowohl das Ergebnis des *Homogenitätsindex* als auch den *Cronbachs Alpha* zeigten gute Ergebnisse, die für die Effektivität der Items sprechen.

#### 6.5.1.3 Evaluation der externen Variable

Die Ergebnisse der Einschätzung der DGS-Kompetenz der Kinder durch die Lehrer und die von den Kindern erreichten Rohwerte zeigten eine signifikante Beziehung. Problematisch aber ist die Variabilität der Selbsteinschätzung der DGS-Kompetenz der Lehrer wie auch deren durchschnittlicher Wert der DGS-Kompetenz. Andere Studien (Herman und Roy, 2006; Johnston, 2004) kommen in dieser Hinsicht zu andersgearteten Ergebnissen. Die Studie von Herman und Roy (2006) ergab auch einen Zusammenhang zwischen der Einschätzung der Tester ($N = 3$) und den Rohwerten der Kinder bei der Durchführung des BSL RST. Johnston (2004) hingegen kommt bei der Adaption des BSL RST in Auslan zu anderen Ergebnissen: Die informelle Einschätzung der Lehrpersonen in der Schule stimmt nicht mit den Rohwerten überein. Für zukünftige Studien ist es daher wichtig, neben einer möglichen Einschätzung der Gebärdensprachkompetenz der Kinder auch valide und reliable Informationen über die Gebärdensprachkompetenz der Lehrpersonen zu erheben. Deswegen ist es sinnvoll, die benutzte Selbsteinschätzungsskala für die DGS zu überarbeiten und für zukünftige wissenschaftliche Studien zu standardisieren. Des Weiteren ist es sinnvoll, in der Zukunft möglichst nur Lehrpersonen, bei denen bekannt ist, dass sie eine hohe DGS-Kompetenz haben, für die Einschätzung der Kinder einzusetzen.

#### 6.5.1.4 Inhaltsvalidität

Die Auswertung der Literatur zum Erwerb von Gebärdensprachen und DGS-Strukturen hat ergeben, dass (1) es in der DGS mit der BSL vergleichbare Strukturen gibt und (2) die Studien zum Gebärdenspracherwerb anderer Gebärdensprachen einen brauchbaren ersten Anhaltspunkt über die Entwicklung bestimmter Strukturen bieten. Deswegen wird hier argumentiert, dass der inhaltlichen Validität des adaptierten DGS-Tests aufgrund der Literaturauswertung Genüge getan wurde.

#### 6.5.1.5 Evaluation der Rohwerte in Beziehung zu anderen Variablen

Die Variablen Alter des Kindes beim Zugang zu einer Gebärdensprache, Hörstatus der Eltern und Lebensalter wurden mit den Rohwerten verglichen, um Unterschiede beziehungsweise zusätzliche Informationen zu erhalten, die die Variabilität in den Rohwerten erklären können.

Die Variable *Alter des Kindes beim Zugang zu einer Gebärdensprache* zeigte, dass Kinder mit einem frühen Zugang zu einer Gebärdensprache (0–3 Jahre alt) bessere Rohwerte erzielten als Kinder mit einem späten Zugang (3–6 Jahre alt). Der frühe Zugang hat einen Einfluss, bietet aber keine kausale Erklärung. Des Weiteren ist die Variable des Zugangs nicht unabhängig vom Hörstatus der Eltern: 21 der 27 Kinder in der Gruppe mit frühem Zugang zu DGS hatten gehörlose Eltern. Der frühe Zugang zu einer Sprache ist eine wichtige Variable für einen erfolgreichen Erstspracherwerb (Mayberry et al., 2002).

Der *Hörstatus der Eltern* zeigte auch einen klaren Bezug zu den Rohwerten: Gehörlose Kinder aus gehörlosen Familien schnitten besser ab als Kinder aus hörenden Familien. Aber auch dies zeigt nur eine Beziehung und keine Kausalität, da zum Beispiel der frühe Zugang eine wichtige Variable ist. Da zu wenige Informationen über das Gebärdensprachalter vorhanden waren, war es leider nicht möglich, dieser Frage weiter nachzugehen. Gehörlose Kinder gehörloser Elter scheinen auch bereits mit 6 bis 7 Jahren die höchsten Rohwerte innerhalb der Gruppe zu erzielen, was möglicherweise darauf hindeutet, dass der Test nicht genügend differenziert für Kinder, die älter als 6 bis 7 Jahre alt sind. Hier sollten zusätzliche Testitems entwickelt werden, die zwischen jüngeren und älteren Kindern gezielter differenzieren können.

Die Ergebnisse in Bezug auf *Lebensalter* und Rohwerte ließen auch klare Schlüsse zu: je älter die Kinder, desto besser die Rohwerte. Auch hier bedeutet dies, dass der Test ein wichtiges Kriterium erfüllt. In Altergruppen eingeteilt (beide Untergruppen zusammen) ergibt sich, dass es keine Korrelation zwischen den Rohwerten und Kindern > 8 Jahre alt gibt. Dies bestätigt die Annahme, dass der Test möglicherweise nicht gut differenzierende Testitems für ältere Kinder dieser Stichprobe beinhaltet.

Diese Ergebnisse sind wichtig zur Definition der Referenzgruppe für die Standardisierung des adaptierten DGS-Tests. Referenzgruppe (gebärdensprachkompetente Kinder) und Zielgruppe (Kinder, die späteren Zugang zu DGS haben, als auch DGS-kompetente Kinder) sind nicht identisch bei der Entwicklung von Gebärdensprachtests.

#### 6.5.1.6 Möglichkeiten der Differenzierung zwischen den Kindern

Der adaptierte DGS-Test sollte Testitems beinhalten, die zwischen älteren und jüngeren Kindern differenzieren. Damit Testitems differenzieren können, sind zwei Themen sehr wichtig: *Komplexität der Testitems* und *Häufigkeit von bestimmten sprachlichen Strukturen*.

*(1) Komplexität von Testitems*: Einige Testitems, die räumliche Konzepte wie *hinter*, *vor* oder *rechts/links* abbilden, sollten aufgrund der Ergebnisse der Itemanalyse entfernt oder überarbeitet werden (Testitems 14, 31, 35, 36, 37, 41 und 44). Eine mögliche Erklärung ist, dass diese Testitems sprachliche Konzepte darstellen, die erst im Alter von 11 bis 12 Jahren erworben werden (Morgan et al., 2008; Slobin et al., 2003). Diese Strukturen sind komplexer als andere und wurden von den Kindern in dieser Studie recht selten richtig beantwortet, da sie diese höchstwahrscheinlich noch nicht erworben hatten (das älteste Kind der Stichprobe war 10;10). Diese Art der Items wäre daher eine gute Möglichkeit der Differenzierung zwischen jüngeren und älteren Kindern, das heißt diese Items sollten überarbeitet und bei einer Standardisierung einbezogen werden. Es sollten auch Kinder, die bis 11 oder 12 Jahre alt sind, in diese Studie einbezogen werden.

*(2) Häufigkeit sprachlicher Strukturen:* Die Häufigkeit des Vorkommens bestimmter sprachlicher Strukturen und deren Abbildung in Testitems kann auch zu einer Differenzierung beitragen. Es ist aus dem englischen Spracherwerb bekannt, dass sprachliche Strukturen, die häufiger vorkommen, vor weniger häufig vorkommenden Strukturen erworben werden (Tomasello, 2003). Der Stand der Forschung über DGS lässt noch keine Aussage über

dieses Thema zu, aber mit dem geplanten 15-jähigen DGS-Korpusprojekt an der Universität Hamburg lassen sich in Zukunft solche Fragen für zukünftige Testentwicklungen und -adaptionen eher beantworten.

#### 6.5.1.7 Methodische Erkenntnisse für die Testadaption

In diesem Abschnitt sollen methodische Erkenntnisse, die für dieses als auch weitere Projekte von Testadaptionen wichtig sind, diskutiert werden. Die folgenden vier Abschnitte behandeln linguistische Themen.

*Regionale Varianten*: Im Rahmen der Testadaption wurden auch regionale Varianten erhoben, die allerdings nicht eindeutig einer Region zugeordnet werden konnten (3.1.2 *Pilot 1 to establish suitability of test items*). Wenn mehr Forschungsergebnisse über regionale Varianten verfügbar sein werden, wird es einfacher werden, auch dieses Problem bei Testentwicklungen einzubeziehen. Eine grundsätzliche Frage bleibt, ob mehrere Versionen eines Tests entwickelt werden sollen oder ob es nicht ausreichend ist, eine Übungseinheit einzubauen, bei der den Kindern bestimmte im Test benutzte Varianten gezeigt werden, wie es im Rahmen dieser Studie umgesetzt wurde. Eine solche Vorgangsweise ist allerdings auch nur dann möglich, wenn es sich nicht um zu viele Gebärden handelt, da sonst die Merkfähigkeit der Kinder überbeansprucht wird.

*Methodische Probleme in der Gebärdensprachforschung*: Ein methodisches Problem für die Testadaption ist, dass die Benutzung unterschiedlicher Modelle/Theorien, die mehr oder weniger das gleiche sprachliche Phänomen beschreiben, sprachübergreifende Studien erschwert (Schembri, 2003). Die Anwendung eines ähnlichen Modells würde mehr sprachvergleichende Forschung ermöglichen und auch einen Vorteil für Testadaptionen bringen.

In einer Erwerbsstudie von Auslan (de Beuzeville, 2006) benutzte die Wissenschaftlerin die gleichen Elizitationsmaterialien wie in einer Erwerbsstudie zu ASL (Schick, 1987). Kinder in der australischen Studie erwarben zum Beispiel *handling classifiers* und *SASS* früher als Kinder in der amerikanischen Studie. Das Beherrschen (*mastery*) der sprachlichen Strukturen (zum Beispiel *handling classifiers*) war in beiden Studien als das Beherrschen der sprachlichen Form wie sie Erwachsene benutzen definiert worden. Der vermeintlich frühere Erwerb in der Studie von de Beuzeville (2006) lässt sich vor allem dadurch erklären, dass in der Studie mehrere Kriterien/Optionen zum Beherrschen der Erwachsenform im Vergleich zur Studie von Schick (1987) akzeptiert wurden. Auch wenn sich Unterschiede in Spracherwerbsstudien durch die benutzte Methodik erklären lassen, so sind diese Ergebnisse vorsichtig abzuwägen bei der Erstellung beziehungsweise Übertragung von Meilensteinen der Entwicklung als Grundlage zur Formulierung von Hypothesen bei der Adaption von Gebärdensprachtests.

*Sprachproduktion und -verständnis*: Die ausgewerteten Spracherwerbsstudien von Gebärdensprachen ergaben eine Mehrheit an Studien, die den Erwerb von Sprachproduktion untersuchten. Nur wenige Studien untersuchten das Sprachverständnis (ASL: Bellugi et al., 1988; BSL: Morgan et al., 2002; Morgan und Woll, 2002b). Basierend auf der Annahme, dass Verständnis vor Produktion erworben wird (Hirsh-Pasek und Golinkoff, 1996; Morgan und Woll, 2002b, 2003), bieten Studien über den Erwerb von Gebärdensprachproduktion bereits einen ersten Hinweis auf bestimmte sprachliche Strukturen, die in Items eines Sprachverständnistests abgebildet sind. Jedoch ist es nicht klar, welche sprachliche Struktur zu welchem Zeitpunkt rezeptiv erworben wird, was aber für die Adaption und Entwicklung von Testitems eine wichtige Erkentnis darstellt. Das bedeutet, dass es keinen Sinn machen würde, rezeptive sprachliche Strukturen in einem Test abzubilden, wenn diese bereits von der Alterszielgruppe erworben sind. Mehr Forschung über den Erwerb von Sprachverständnis könnte dazu beitragen, adäquatere Testitems zu entwickeln.

*Sprachspezifische Strukturen*: Ein Mangel an Forschungsergebnissen über sprachspezifische Strukturen stellt ein Hindernis in der Adaption und Entwicklung von Gebärdensprachtests dar. Es gibt zu wenige typologisch motivierte Studien im größeren Rahmen (zum Beispiel Zeshan, 2006), die es zulassen würden, mehr Wissen über Unterschiede und Ähnlichkeiten zwischen Gebärdensprachen zur Verfügung zu haben, was eine wichtige Voraussetzung für eine erfolgreiche Testadaption ist.

### 6.5.2 Konsequenzen in Bezug auf die Standardisierung

Eine Standardisierung des adaptierten DGS-Tests zur Verwendung in den Schulen sollte durchgeführt werden.

Die *Größe der Stichprobe* in dieser Studie ($N = 54$) ist aus statistischer Sicht betrachtet nicht ausreichend genug, um repräsentative Aussagen machen und komplexere statistische Modelle (zum Beispiel für das Gebärdensprachalter) rechnen zu können. Dies ist in diesem Bereich immer schwierig, da die Population sehr klein ist. Eine genaue Normierungsstichprobe zu definieren ist ein komplexes Unterfangen, da es zu wenige Informationen über die Gesamtpopulation „hörgeschädigter Kinder“ gibt. Daher werden – in Anlehnung an die Studie zur Entwicklung des BSL RST (Herman et al., 1998) – zwei Kriterien für eine Normierung des adaptierten DGS-Tests definiert: (1) qualitatives Kriterium zum sprachlichen Hintergrund der Kinder: DGS-kompetente gehörlose und hörende Kinder gehörloser und hörender Eltern (wie beim BSL RST) und (2) Definition von 6 Altergruppen (3;0–3;11, 4;0–4;11, 5:0–5;11, 6;0–7;11, 8;0–9;11 und 10;0+) mit mindestens 30 Kindern in jeder Altergruppe, das heißt eine (Teil)Normierungsstichprobe von mindestens 180 Kindern ingesamt.

*Elternfragebogen*: Um in der Standardisierungsstudie die Informationen der Eltern verwenden zu können, sollte der Fragebogen zumindest in die wichtigsten Minderheitensprachen übersetzt werden, zum Beispiel DGS und Türkisch.

*Schülerfragebogen*: Die Skala der Einschätzung der DGS-Kompetenz der Schüler durch die Lehrer sollte abgeändert werden von 1 bis 10 (1 = Minimum, 10 = Maximum), um Missverständnissen in der Standardisierungsstudie vorzubeugen. Außerdem sollten möglichst nur Lehrer mit hoher DGS-Kompetenz diese Information eintragen (zum Beispiel gehörlose Lehrer).

*Lehrerfragebogen*: Die Skala der Selbsteinschätzung der DGS-Kompetenz der Lehrer sollte überarbeitet werden für die Standardisierung.

*Testen mit jüngeren Kindern*: Nicht alle jüngeren Kinder waren in der Lage, eine Maus zu benutzen, daher sollte eine kleine Voruntersuchung mit einer speziellen „Kinder-Maus“ noch mehr Aufschluss bringen. Der Test könnte andererseits aber auch auf einen Laptop mit Touchscreen-Technologie übertragen werden.

*Gebärdensprachalter*: Der Effekt des Gebärdensprachalters auf die Testergebnisse konnte leider nicht ermittelt werden, da zu wenige Daten zur Verfügung standen. Mit übersetzten Elternfragebögen könnten möglicherweise mehr valide Information über das Gebärdensprachalter erhoben werden.

*Validität*: Um die Validität zu erhöhen, sollte versucht werden, den Test von (gehörlosen) Linguisten begutachten zu lassen. Leider war dies im Rahmen der vorliegenden Studie aufgrund der vorhandenen begrenzten Ressourcen nicht möglich. Des Weiteren sollte neben diesem hier adaptierten DGS-Verständnistest auch der Perlesko, ein Verfahren zur Überprüfung der (rezeptiven) lexikalisch-semantischen Fähigkeiten in der DGS (Bizer und Karl, 2002) durchgeführt werden, um die Übereinstimmungsvalidität zu ermitteln.

### 6.5.3 Empirische Erkenntnisse für zukünftige Testadaptionen

Es wären sowohl mehr DGS-Erwerbsstudien wichtig (nicht nur zur Produktion, sondern auch zum Verständnis) als auch mehr sprachvergleichende Studien zwischen Gebärdensprachen (unter Einbezug von gesprochenen Sprachen).

*Voruntersuchung 2 mit gehörlosen Erwachsenen*: Für eine zukünftige Voruntersuchung mit gehörlosen Erwachsenen sollte ein offener Fragebogen benutzt werden, um noch mehr qualitative Daten zu erheben, die möglicherweise zu einer Testüberarbeitung beitragen können.

### 6.5.4 Theoretische Erkenntnisse: Modell zur Testadaption

Die Erkenntnisse der Adaption des DGS-Verständnistests wurden systematisiert und führten zu dem folgenden Modell zur Adaption von Gebärdensprachtests. Es wurde bereits im Literaturrückblick argumentiert, für die Übertragung von einem Ausgangs- in einen Zieltest das Modell der Adaption zu verwenden (van de Vijver und Poortinga, 2005). Das Modell, das hier vorgeschlagen werden wird, beinhaltet verschiedene methodische und theoretische Schritte, die im Anschluss zur Konstruktdefinition und -validierung im Modell zur Testadaption zusammengefasst und im Appendix J-2 genau aufgelistet sind. Ein wichtiger theoretischer Beitrag zu diesem Modell ist die Konstruktdefinition als auch ein vorgeschlagenes methodisches Vorgehen, das Konstrukt durch eine zusätzliche externe Variable (Einschätzung durch Gehörlose) zu validieren.

#### 6.5.4.1 Annäherung an eine Konstruktdefinition

Das Konstrukt eines Tests muss zuerst definiert werden (van de Vijver und Leung, 1997a), auch wenn angenommen werden kann, dass der BSL RST (Herman et al., 1999) und der adaptierte DGS-Verständnistest das gleiche Konstrukt testen (das heißt *Sprachentwicklung*). Ein Konstrukt kann als eine Fähigkeit oder Fähigkeiten definiert werden, die sich in den Testergebnissen widerspiegelt beziehungsweise widerspiegeln (Davies et al., 1999). Es ist wichtig, während der Entwicklung eines Tests zu entscheiden, was genau mit diesem Test getestet werden soll. Die genaue Definition der Fähigkeiten, die mit einem Sprachtest erfasst werden sollen, ist Gegenstand der Konstruktdefinition (Bachman, 1990; Douglas, 2000). Die genaue Beschreibung und Benennung der sprachlichen Fähigkeiten kann auf unterschiedlichsten Ebenen spezifiziert werden, zum Beispiel als *grammatikalische Kompetenz* einer Sprache oder etwas spezifizierter als *Morphologie* und *Syntax* (Bachman, 1990). Im Rahmen der vorliegenden Studie soll das Konstrukt basierend auf der Literaturauswertung von Studien zu Gebärdensprachen weiter spezifiziert werden.

Das Konstrukt für den adaptierten DGS-Test kann als die *Entwicklung von Morphologie und Syntax* definiert werden. Dies beinhaltet einfache und komplexere morphologische und syntaktische Strukturen der DGS, die wiederum mit dem Alter zusammenhängen und so den Entwicklungsaspekt abdecken. Die morphologischen und syntaktischen Strukturen können nochmals genauer spezifiziert werden als die sprachlichen Bereiche, die in dem Test abgebildet werden (Figure 5.1).

#### 6.5.4.2 Operationalisierung des Konstrukts

In einem nächsten Schritt muss das Konstrukt operationalisiert werden, das heißt das Konstrukt muss beobachtbar gemacht werden in Form der Entwicklung oder Adaption von Testitems (Bachman, 1990; Bachman und Palmer, 1996). Ein operationalisiertes Konstrukt muss immer validiert werden.

Ein Schritt in der Konstruktoperationalisierung ist es durch gehörlose Testentwickler die adaptierten Testitems anhand der Abbildung *Map Ranking of Item Complexity* (Figure 2.2) entsprechend ihres zugeschriebenen Erwerbsalters in eine Abfolge/Rangordnung zu bringen. Dieses Vorgehen wird als *Ranking 1 (Operationalisierung)* bezeichnet werden. Die Anordnung der Testitems kann sowohl (1) dem Entwicklungsaspekt als auch (2) den DGS-spezifischen Strukturen Rechnung tragen. Die Rangfolge der Testitems des BSL RST, die entsprechend ihres Schwierigkeitsgrades angeordnet sind, kann hier auch eine zusätzliche Hilfestellung geben. Der Operationalisierung folgen mehrere Schritte von Voruntersuchungen, die ihrerseits in Überarbeitungen des Tests münden (Abbildung 6.2: Modell zur Testadaption).

#### 6.5.4.3 Validierung des Konstrukts

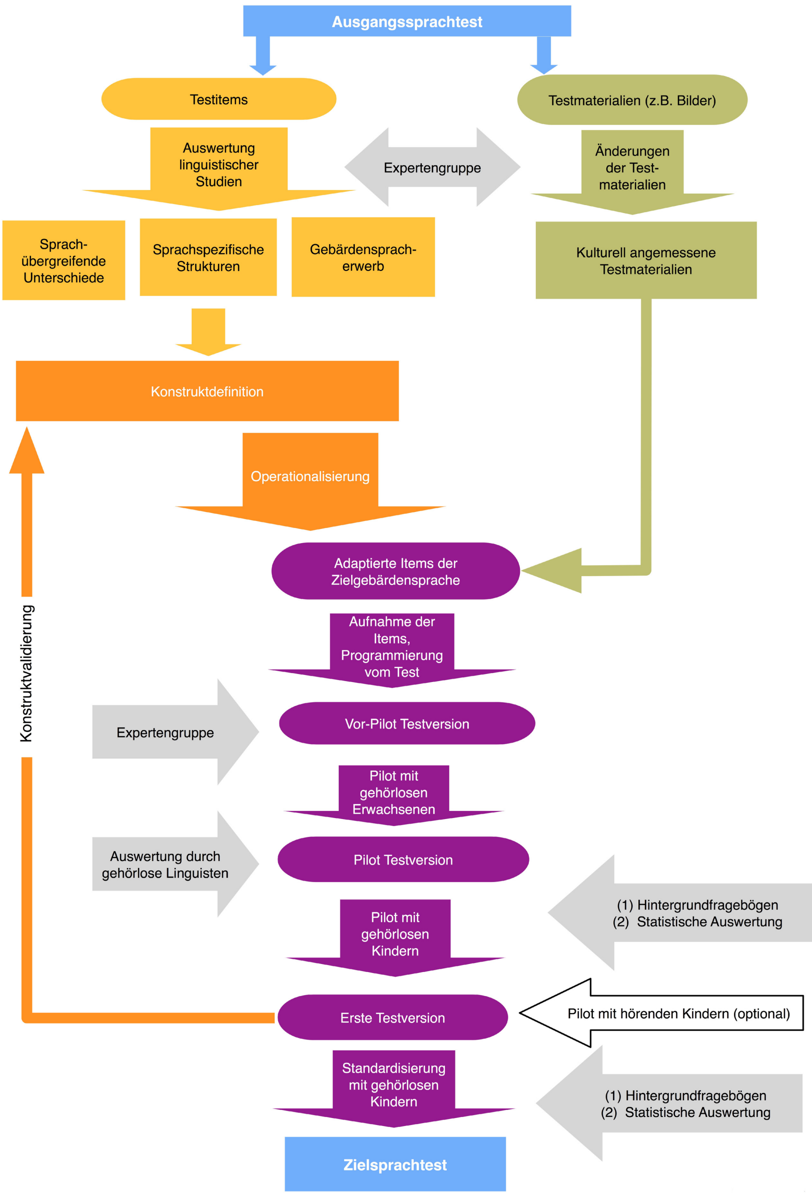
Nach der Operationalisierung des Konstrukts müssen die Testitems, also das operationalisierte Konstrukt, validiert werden. Das hier vorgeschlagene methodische Vorgehen sieht vor, dass gehörlose Erwachsene, die nicht in den Prozess der Testadaption involviert waren, die abgebildeten Testitems (beziehungsweise ihre abgebildeten sprachlichen Strukturen) in Bezug auf das ihnen zugeschriebene relative Erwerbsalter einschätzen. Dieses Vorgehen wird als *Ranking 2 (Validierung)* bezeichnet werden. Die Ergebnisse vom *Ranking 2* werden dann mit den Ergebnissen des ursprünglichen *Ranking 1*, verglichen werden, was einer ersten Überprüfung des operationalisierten Konstrukts gleich kommt. Eine weitere Möglichkeit ist es, das Ergebnis des *Ranking 2* mit den Ergebnissen der Itemanalyse abzugleichen, was ein erster Hinweis zum Schwierigkeitsgrad der Testitems darstellt. Dieser methodische Ansatz, sprachliche Strukturen in Bezug auf ihr relatives Erwerbsalter einzuordnen, wurde in einer Studie zur Einschätzung des Erwerbsalters von lexikalischen Einheiten der BSL erfolgreich genutzt (Vinson et al., 2008).

Die hier formulierten Schritte zur Konstruktdefinition und -validierung können gleichermaßen in zukünftigen Testadaptionen und -entwicklungen angewendet werden.

#### 6.5.4.4 Modell zur Testadaption

Adaption wurde als der bevorzugte Ansatz zur Testübertragung definiert. Neben der Zusammenfassung aller methodischen Schritte ist der Hauptbeitrag von diesem Modell die Konstruktdefinition und die vorgeschlagene Methode zur Überprüfung der Validität des Konstrukts. Die verschiedenen Schritte im Prozess der Testadaption sind in dem Modell (Abbildung 6.2) zusammengefasst und in dem Appendix J-2 übersichtlich dargestellt.

Abbildung 6.2: Modell zur Testadaption



### 6.5.5 Zusammenfassung und Schlussfolgerung

Der Hauptbeitrag dieser Studie ist es, kulturelle, linguistische, methodische und theoretische Themen und Überlegungen im Prozess der Testadaption an unterschiedlichen Phasen der Testadaption darzustellen. Die theoretischen Erkenntnisse ergaben die Konstruktdefinition, Ansätze zur Validierung des Konstrukts so wie den Vorschlag eines Adaptionsmodells, das methodische, kulturelle und linguistische Themen abdeckt. Dieses Modell kann gleichermaßen für Testentwicklung oder -adaption verwendet werden.

Auf einer konkreten Ebene bieten die Ergebnisse dieser Studie eine gute Grundlage für die Standardisierung des DGS-Verständnistests. Die Anwendung von neuer Technologie in diesem Bereich zu Testzwecken stellt auch einen neuen und vielversprechenden Ansatz dar.

Die ausgewerteten Erwerbsstudien zu Gebärdensprachen bieten eine wichtige Grundlage zu einer theoretischen Diskussion von sprachvergleichenden Studien. Auf der einen Seite kann die zunehmende Anzahl an Erwerbsstudien über Gebärdensprachen als eine wichtige Grundlage für Testentwicklung und -adaption dienen. Auf der anderen Seite können Daten, die in einem größeren Umfang durch Testentwicklung oder -adaption gewonnen werden, zu einem besseren Verständnis von Gebärdensprachentwicklung beitragen. Der Zuwachs an Wissen über Gebärdenspracherwerb und Testentwicklung beziehungsweise -adaption ist gegenseitig gegeben und kann so zu einer Weiterentwicklung in dem jeweiligen Bereich beitragen.

# Bibliography


Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation* (9th printing, 2005). Cambridge: Cambridge University Press.

Alant, E., & Beukes, S. M. (1986). The application of the Peabody Picture Vocabulary Test – Revised (PPVT-R) to non-mainstream children. *The South African Journal of Communication Disorders, 33*, 7–14.

Anderson. D., & Reilly, J. (1997). The puzzle of negation: How children move from communicative to grammatical negation in ASL. *Applied Psycholinguistics*, *18*, 411–429.

Anderson, D., & Reilly, J. (2002). The MacArthur Communicative Development Inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education, 7*(2), 83–119.

Audeoud, M., & Haug, T. (2008). "Grundsätzlich wollen wir Tests, die alle sprachlichen Ebenen überprüfen!" Eine Pilot-Studie zum Bedarf an Gebärdensprachtests für hörgeschädigte Kinder an Deutschschweizer Hörgeschädigtenschulen ["We want assessment instruments, that test all areas of a language!" A pilot survey investigating the need for sign language tests for deaf children at Swiss-German schools for the deaf]. *Hörgeschädigtenpädagogik, 61*(1), 15–20.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baker, A., & Jansma, S. (2005). *NGT-OP Een Screening Instrument voor Peuters in de Nederlandse Gebarentaal em Handleiding* [NGT-OP a screening instrument for young children in NGT and handbook]. Amsterdam: University of Amsterdam Internal Publication.

Baker, A., van den Bogaerde, B., & Woll, B. (2008). Methods and procedures in sign language acquisition studies. In A. Baker & B. Woll (Eds.), *Sign language acquisition* (pp. 1–49). Amsterdam: John Benjamins Publishing Company.

Bartram, D. (1990). Introduction. In J. R. Beech & L. Harding (Eds.), *Testing people: A practical guide to psychometric*s (pp. 1–10). Windsor: NFER Nelson.

Bartram, D. (2006). Testing on the Internet: Issues, challenges, and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the Internet – Issues and advances* (pp.13–38). Chichester, West Sussex: John Wiley & Sons, Ltd.

Bates, E. (1993). Comprehension and production in early language development: Comments on Savage-Rumbaught et al. *Monographs of the Society for Research in Child Development*, *Serial No. 233*, *58*(3–4) (pp. 222–242). Retrieved on July 19, 2008 from http://crl.ucsd.edu/~bates/papers/pdf/from-meiti/47-savage-rumbaugh.comments.pdf.

Becker, C. (2003). *Verfahren der Lexikonerweiterung in der Deutschen Gebärdensprache* [Methods of lexical growth in German Sign Language]. Hamburg, Germany: Signum Verlag.

Beecken, A., Keller, J., Prillwitz, S., & Zienert, H. (1999). *Grundkurs Deutsche Gebärdensprache – Stufe I Lehrbuch* [Introductory course to German Sign Language]. Hamburg, Germany: Signum Verlag.

Bellugi, U., van Hoek, K., Lillo-Martin, D., & O'Grady, L. (1988). The acquisition of syntax and space in young deaf signers. In K. Mogford-Bevan & D. Bishop (Eds.), *Language development in exceptional circumstances* (pp. 132–149). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bernardino, E. (2005). *The acquisition of classifiers in verbs of motion and verbs of location in Brazilian Sign Language.* Unpublished doctoral dissertation, Boston University, Graduate School of Arts and Sciences (ProQuest UMI number: 3186488).

Bizer, S., & Karl, A.-K. (2002). *Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schrift- und Lautsprache* [Development of a vocabulary test for deaf elementary school-aged children for sign, written, and spoken language]. Unpublished doctoral dissertation, School of Education, Hamburg University.

Bornstein, M. H., Haynes, O. M., Painter, K. M., & Genevro, J. L. (2000). Child language with mother and with stranger at home and in the laboratory: A methodological study. *Journal of Child Language, 27*, 407–420.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage) [Statistics for social scientists (5th ed.)]. Heidelberg, Germany: Springer.

Bortz, J., & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Auflage) [Research methods and evaluation for humanities and social sciences (3rd ed.)]. Heidelberg, Germany: Springer.

Bos, H. (1994). An auxiliary verb in Sign Language of the Netherlands. In I. Ahlgren, B. Bergman, & M. Brennan (Eds.), *Perspectives on sign language structure: Vol. 1* (pp. 37–53). Durham, UK: ISLA.

Boudreault, P. (1999). *Grammatical processing in American Sign Language: Effects of age of acquisition and syntactic complexity*. Unpublished master thesis, School of Communication Sciences and Disorders, McGill University, Montreal, Canada.

Boudreault, P., & Mayberry, R. I. (2000). *Grammatical processing in American Sign Language: Effects of age of acquisition and syntactic complexity*. Poster presented at the 7th International Conference on Theoretical Issues in Sign Language Research, Amsterdam, The Netherlands, July 23–27, 2000.

Bowerman, M. (1996). Learning how to structure space for language: A cross-linguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garett (Eds.), *Language and space* (pp. 385–436). Cambridge, MA: The MIT Press.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Genter & S. Goldin-Meadow (Eds.), *Language in mind: Advance in the study of language and thought* (pp. 387–427). Cambridge, MA: The MIT Press.

Boyes Braem, P. (1994). Acquisition of the handshape in American Sign Language: A preliminary analysis. In V. Volterra & C. J. Erting (Eds.), *From gesture to language in hearing and deaf children* (pp. 107–127). Washington, DC: Gallaudet University Press.

Boyes Braem, P. (1995). *Einführung in die Gebärdensprache und ihre Erforschung* [Introduction into sign language and its investigation]. Hamburg, Germany: Signum Verlag.

Blamey, P. J. (2003). Development of spoken language by deaf children. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 232– 246). Oxford: Oxford Univer-sity Press.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality and Social Psychology, 54*(1), 106–148.

Brooks, P. J. (2004). Grammatical competence is not a psychologically valid construct. *Journal of Child Language, 31*, 467–470.

Brown, H. D. (2004). *Language assessment – Principles and classroom practices*. White Plains, NY: Pearson Education.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Auflage) [Introduction into the construction of tests and questionnaires (2nd ed.)]. München, Germany: Pearson Studium.

Camarata, S., & Nelson, K. E. (2002). Measurement and the diagnosis and treatment of language disorders in children. *Peabody Journal of Education, 77*(2), 106–116.

Canale, M., & Swain, M. (1980). Theoretical based communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*, 69–111.

Chapelle, C. A. (1998). Construct definition and validity inquiry. In L. F. Bachman & A. C. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge: Cambridge University Press.

Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, *19*, 254–272.

Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology.* Cambridge: Cambridge University Press.

Chavez, E. L. (1982). Analysis of a Spanish translation of the Peabody Picture Vocabulary Test. *Perceptual and Motor Skills*, *54*, 1335–1338.

Cheung, F. M. (1985). Cross-cultural considerations for the translation and adaptation of the Chinese MMPI in Hong Kong. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment: Vol. IV* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum Associates.

Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, *20*(3), 295–320.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–121.

Christensen, K. M., & Delgado, G. L. (Eds.). (1993). *Multicultural issues in deafness.* White Plains, NY: Longman Publishing Group.

Church, A. T. (1987). Personality research in a non-Western culture: The Philippines. *Psychological Bulletin*, *102*(2), 272–292.

Clark, E. V. (2003). *First language acquisition.* Cambridge: Cambridge University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing – Studies in language testing 7*. Cambridge: Cambridge University Press.

De Beuzeville, L. (2004). The acquisition of classifier signs in Auslan (Australian Sign Language) by deaf children from deaf families: A preliminary analysis. *Deaf Worlds*, *20*(2), 120–140.

De Beuzeville, L. (2006). *Visual and linguistic representation in the acquisition of depicting verbs: A study of native signing deaf children of Auslan (Australian Sign Language).* Unpublished doctoral dissertation, Renwick College, University of Newcastle, Australia.

Debevc, M., & Peljhan, Z. (2004). The role of video technology in on-line lectures for the deaf. *Disability and Rehabilitation*, *26*(17), 1048–1059.

Diehl, J. M., & Staufenbiel, T. (2002). *Statistik mit SPSS – Version 10 & 11* [Statistics with SPSS – Version 10 & 11]. Frankfurt/Main, Germany: Verlag Dietmar Klotz.

Douglas, D. (2000). *Assessing language for specific purposes.* Cambridge: Cambridge University Press.

Dubuisson, C., Parisot, A.-M., & Vercaingne-Ménard, A. (2008). Bilingualism and deafness – Correlation between deaf students' ability to use space in Quebec Sign Language and their reading comprehension in French. In C. Plaza-Pust & E. Morlas-López (Eds.), *Sign bilingualism – Language development, interaction, and maintenance in sign language contact situations* (pp. 51–71). Amsterdam: John Benjamins Publishing Company.

Dunn, L. M., & Dunn, L. M. (2006). *Peabody Picture Vocabulary Test – 4*. Bloomington, MN: Pearson Assessment.

Enns, C., & Zimmer, K. (2009). *Research study: Adapting the British Sign Language Receptive Skills Test into American Sign Language. Summary Report*. Retrieved on June 10, 2010 from http://home.cc.umanitoba.ca/~ennscj/ASLtestsummary.pdf.

Fehrmann, G., Huber, W., Jäger, L., Sieprath, H., & Werth, I. (1995a). *Linguistische Konzeption des Aachener Tests zur Basiskompetenz in Deutscher Gebärdensprache (ATG)* [Linguistic conception of the Aachen test of basic competence of German Sign Language (ATG)]. Unpublished manuscript, RWTH-Aachen, Germanistisches Institut & Neurologische Klinik, Projekt DESIRE.

Fehrmann, G., Huber, W., Jäger, L., Sieprath, H., & Werth, I. (1995b). *Aufbau des Aachener Tests zur Basiskompetenz in Deutscher Gebärdensprache (ATG)* [Design of the Aachen test of basic competence of German Sign Language (ATG)]. Unpublished manuscript, RWTH-Aachen, Germanitisches Institut & Neurologische Klinik, Projekt DESIRE.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* San Diego, CA: Singular Publishing Group.

Fish, S., Morén, B., Hoffmeister, R., & Schick, B. (2003). The acquisition of classifier phonology in ASL by deaf children: Evidence from descriptions of objects in specific spatial arrangements. In B. Beachley, A. Brown, & F. Conlins (Eds.), *Proceedings of the 27th annual Boston University Conference on Language Development* (pp. 252–263). Somerville, MA: Cascadilla Press.

Fischer, S. (1998). Critical periods for language acquisition: Consequences for deaf education. In A. Weisel (Ed.), *Issues unresolved: New perspectives on language and deaf education* (pp. 9–26). Washington, DC: Gallaudet University Press.

Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik – Mit Hinweisen zur Intervention* (3. Auflage) [Textbook for psychological diagnosis – With references to intervention (3rd ed.)]. Göttingen, Germany: Hogrefe Verlag.

Friend, M., & Keplinger, M. (2008). Reliability and validity of the computerized comprehension task (CCT): Data from American English and Mexican Spanish infants. *Journal of Child Language, 35*, 77–98.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment – An advanced resource book.* London: Routledge Applied Linguistics.

Gard, A., Gilman, L., & Gorman, J. (1993). *Speech and language development chart* (2nd ed.). Austin, TX: PRO-ED, Inc.

Gaur, A. S., & Gaur, S. S. (2006). *Statistical methods for practice and research – A guide to data analysis using SPSS*. New Delhi: Response Books.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*, 304–312.

Genter, D., & Bowerman, M. (2008). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480). London: Psychology Press.

Gerken, L., & Shady, M. E. (1996). The picture selection task. In D. McDaniel, C. McKee, & H. Smith Cairns (Eds.), *Methods for assessing children's syntax* (pp. 125–146). Cambridge, MA: The MIT Press.

Gerner de Garcia, B. (2000). Meeting the needs of Hispano/Latino deaf students. In K. M. Christensen & G. L. Deldago (Eds.), *Deaf plus. A multicultural perspective* (pp. 149–198). San Diego, CA: Dawn Sign Press.

Glück, S. (2001). Morphosyntaktische Eigenschaften der Klassifikation in Deutscher Gebärdensprache [Morphosyntactic characteristics on classification in German Sign Language]. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000 – Theorie und Anwendung* [Sign language linguistics 2000 – Theory and application] (pp. 127–145) Hamburg, Germany: Signum Verlag.

Glück, S. (2005). Dimensionen – Klassifikation und Spezifizierung in Gebärdensprachen und Lautsprachen [Dimensions – Classification and specification in sign and spoken languages]. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* [Sign languages: Structure, acquisition, use] (pp. 179–200). Hamburg, Germany: Helmut Buske Verlag.

Glück, S., & Pfau, R. (1997a). Eine Klasse für sich: Klassifizierende Verben in Deutscher Gebärdensprache [A class of its own: Classifying verbs in German Sign Language]. *Zeitschrift für Sprachwissenschaften*, *16*(1/2), 181–208.

Glück, S., & Pfau, R. (1997b). Einige Aspekte der Morphologie und Morphosyntax in Deutscher Gebärdensprache [Some aspects on the morphology and morphosyntax of German Sign Language]. *Frankfurter Linguistische Forschung*, *20*, 30–48.

Glück, S., & Pfau, R. (1998). On classifying classification as a class of inflection in German Sign Language. In T. Cambier-Langeveld, A. Lipták, & M. Redford (Eds.), *Proceedings of ConSOLE VI* (pp. 59–74). Leiden, The Netherlands: Sole.

Glück, S., & Pfau, R. (1999). A distributed morphology account of verbal inflection in German Sign Language. In T. Cambier-Langeveld, A. Lipták, M. Redford, & E. J. van de Torre (Eds.), *Proceedings of ConSOLE VII* (pp. 60–85). Leiden, The Netherlands: Sole.

Goldin-Meadow, S., & Mylander, C. (1990). Beyond the input given: The child's role in the acquisition of language. *Language, 66*(2), 323–355.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*, 279–281.

Grosjean, F. (2008). *Studying bilinguals.* Oxford: Oxford University Press.

Große, K.-D. (2003). *Das Bildungswesen für Hörbehinderte in der Bundesrepublik Deutschland – Daten und Fakten zu Realitäten und Erfordernissen* [The educational system for the hearing-impaired in the Federal Republic of Germany]. Heidelberg, Germany: Universitätsverlag Winter.

Große, K.-D. (Ed.). (2004). *Hörbehinderte Schülerinnen und Schüler unterschiedlicher nationaler Herkunft – Eine internationale Herausforderung an die Hörgeschädigtenpädagogik* [Hearing impaired students of different national background – An international challenge for the education of the hearing impaired]. Heidelberg, Germany: Universitätsverlag Winter.

Günther, K. B. (1999). *Bilingualer Unterricht mit gehörlosen Grundschülern – Zwischenberich zum Hamburger Bilingualen Schulversuch* [Bilingual instruction with deaf elementary school students]. Hamburg, Germany: Verlag hörgeschädigter Kinder.

Günther, K. B., & Hennies, J. (Eds.). (in press). *Bilingualer Unterricht in Gebärden-, Laut- und Schriftsprache mit gehörlosen SchülerInnen in der Primarstufe: Zwischenbericht zum Berliner Bilingualen Schulversuch* [Bilingual education in sign language, spoken language, and written language for deaf primary school students: First report on the bilingual trial class of Berlin]. Berlin: Druckerei der Humboldt-Universität.

Günther, K. B., Hennies, J., & Hintermair, M. (2009). Trends and developments in deaf education in Germany. In D. Moores & M. Miller (Eds.), *Deaf people around the world: Educational, developmental, and social perspectives* (pp. 178–193). Washington, DC: Gallaudet University Press.

Günther, K. B., & Schäfke, I. (2004). *Bilinguale Erziehung als Förderkonzept für gehörlose SchülerInnnen – Abschlußbericht zum Hamburger Bilingual Schulversuch* [Bilingual education as an advancement for deaf students – Final report from the Hamburg bilingual pilot class]. Hamburg, Germany: Signum Verlag.

Hänel, B. (2003). *Der Erwerb der Deutschen Gebärdensprache als Erstsprache – Die frühkindliche Sprachentwicklung von Subjekt- und Objektverbkongruenz in DGS.* [The acquisition of German Sign Language as a first language – The early language acquisition of subject and object verb agreement in German Sign Language]. Unpublished doctoral dissertation, Hamburg University, Department of Linguistics.

Hänel, B. (2005). The acquisition of agreement in DGS: Early steps into spatially expressed syntax. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* [Sign languages: Structure, acquisition, and use] (pp. 201–232). Hamburg, Germany: Helmut Buske Verlag.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, *10*(3), 229–244.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*(3), 164–172.

Hambleton, R. K. (2005). Issues, design, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K., & Kanjee, A. (1995). Increasing the validity of cross-cultural assessments: Use of improved methods for test adaptation. *European Journal of Psychological Assessment, 11*, 147–157.

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, *45*, 153–171.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with the British communicative development inventory. *Journal of Child Language*, *27*, 689–705.

Hammill, D., Brown, V., Larsen, S., & Wiedreholt, J. L. (1994). *Test of adolescent and adult language* (3rd ed.). Austin, TX: PRO-ED, Inc.

Happ, D. (2005). Manuelle und nicht manuelle Module der Deutschen Gebärdensprache (DGS): Linguistische Aspekte [Manual and non-manual components in German Sign Language: Linguistic aspects]. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* [Sign languages: Structure, acquisition, use] (pp. 9–28). Hamburg, Germany: Helmut Buske Verlag.

Happ, D., & Vorköper, M.-O. (2005). Einige Bemerkungen zur syntaktischen und morphologischen Repräsentation von Numerus in Deutscher Gebärdensprache [Comments regarding the syntactic and morphological representation of number in German Sign Language]. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* [Sign languages: Structure, acquisition, use] (pp. 87–110). Hamburg, Germany: Helmut Buske Verlag.

Hart, R. A., & Moore, G. T. (2005). The development of spatial cognition: A review. In R. M. Downs & D. Stea (Eds.), *Image & environment – Cognitive mapping and spatial behavior* (pp. 246–288). Piscataway, NJ: Aldine Transaction.

Hatzopoulou, M. (2008). *The acquisition of reference to self and others in Greek Sign Language – From pointing gesture to pronominal pointing sign.* Unpublished doctoral dissertation, University of Stockholm, Sweden. Retrieved on June 17, 2009 from http://www.diva-portal.org.

Haug, T. (2008a). Review of sign language assessment instruments. In A. Baker & B. Woll (Eds.), *Sign language acquisition* (pp. 51–85). Amsterdam: John Benjamins Publishing Company.

Haug, T. (2008b). Adaptation and evaluation of a computer-based receptive skills test for German Sign Language – Report of an ongoing research project. In *Gebärdensprachlinguistik und Gebärdensprachkommunikation – Referate der VERBAL-Sektion "Gebärdensprachlinguistik und -kommunikation", 34. Österreichische Linguistiktagung an der Universität Klagenfurt am 8.12.2006* (pp. 73–84). Klagenfurt, Austria: Veröffentlichungen des Zentrums für Gebärdensprache und Hörbehindertenkommunikation der Universität Klagenfurt, Band 13.

Haug, T., & Hintermair, M. (2003). Ermittlung des Bedarfs von Gebärdensprachtests für gehörlose Kinder – Ergebnisse einer Pilotstudie [Surveying the need for sign language tests for deaf children – Results of a pilot study]. *Das Zeichen*, *64*, 220–229.

Haug, T., & Mann, W. (2007). *Impact of age on German Sign Language performance of deaf students from a diverse cultural and linguistic background (DCLB).* Paper presented at the 6th International Symposium on Bilingualism, Hamburg, Germany, May 30 to June 2, 2007.

Haug, T., & Mann, W. (2008). Adapting tests of sign language assessment to other sign languages – A review of linguistic, cultural, and psychometric problems. *Journal of Deaf Studies and Deaf Education, 13*(1), 138–147.

Hauser, P. C., Supalla, T., & Bavelier, D. (2008). American Sign Language-Sentence Reproducation Test. In R. M. de Quadros (Ed.), *Sign languages: Spinning and unraveling the past, present and future. TISLR 9, forty-five papers and three posters from the 9. Theoretical Issues in Sign Language Research Conference, Florianopolius, Brazil, December 2006* (pp. 160–172). Petrópolis/RJ, Brazil: Editora Arara Azul. Retrieved on June 5, 2009 from http://www.editora-arara-azul.com.br/EstudosSurdos.php.

Haynes, W. O., & McCallion, M. B. (1981) Language comprehension testing: The influence of three modes of test administration and cognitive tempo on the performance of preschool children. *Language, Speech, and Hearing Services in Schools*, *12*, 74–81.

Haynes, W. O., Purcell, E., & Haynes, M. D. (1979). A pragmatic aspect of language sampling. *Language, Speech, and Hearing Services in Schools*, *10*, 104–110.

Herman, R. (1998). The need for an assessment of deaf children's signing skills. *Deafness and Education: Journal of the British Association of the Teachers of the Deaf*, *22*(3), 3–8.

Herman, R. (2002). *Assessment of BSL development*. Unpublished doctoral dissertation, City University London.

Herman, R., Holmes, S., & Woll, B. (1998). *Design and standardisation of an assessment of British Sign Language development for use with young deaf children: Final report, 1998*. Unpublished manuscript, City University London, UK.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL development: Receptive Skills Test*. Coleford, UK: Forest Books.

Herman, R., Holmes, S., Woll, B. (2008). *Adapting the BSL Receptive Skills Test.* Unpublished manuscript, City University London.

Herman, R., Grove, N., Holmes, S., Morgan, G., Sutherland, H., Woll, B. (2004). *Assessing BSL development: Production test (narrative skills).* London: City University Publication.

Herman, R., & Roy, P. (2006). Evidence from the wider use of the BSL Receptive Skills Test. *Deafness and Education International, 8*(1), 33–47.

Hermans, D., Knoors, H., & Verhoeven, L. (2010). Assessment of sign language development: The case of deaf children in the Netherlands. *Journal of Deaf Studies and Deaf Education*, *15*(2), 107–119.

Heßmann, J. (2001a). *GEHÖRLOS SO! Materialien zur Gebärdensprache – Band 1: Grundlagen und Gebärdenverzeichnis*. Hamburg, Germany: Signum Verlag.

Heßmann, J. (2001b). *GEHÖRLOS SO! Materialien zur Gebärdensprache – Band 2: Gebärdensprachtexte*. Hamburg, Germany: Signum Verlag.

Hirsh-Pasek, K., & Golinkoff, R. M. (1991). Language comprehension: A new look at some old themes. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and* *behavioral determinants of language development* (pp. 301–320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. Smith Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). Cambridge, MA: The MIT Press.

Hoffmeister, R. (1992). *Mapping number and arrangement in the language of deaf children: Plurals and many things.* Unpublished manuscript, Center for the Study of Communication & the Deaf, Boston University.

Hoffmeister, R. (1999). *American Sign Language Assessment Instrument (ASLAI).* Unpublished manuscript, Center for the Study of Communication & the Deaf, Boston University.

Hoffmeister, R. (2000). A piece of puzzle: ASL and reading comprehension in deaf children. In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 143–163). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoffmeister, R., Philip, M. J., Costello, P., & Grass, W. (1997). *Evaluating American Sign Language in deaf children: ASL influences on reading with a focus on classifiers, plurals, verbs of motion and location*. Unpublished manuscript, Center for the Study of Communication & the Deaf, Boston University.

Hoiting, N. (2009). *The acquisition of sign language by deaf toddlers.* Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Hoiting, N., & Slobin, D. I. (2002). Transcription as a tool for understanding: The Berkeley Transcription System for sign language research (BTS). In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 55–75). Amsterdam: John Benjamins Publishing Company.

Hooper, S., Rose, S., & Miller, C. (2005). Assessing American Sign Language performance: Developing an environment for capturing, evaluating, and monitoring student progress. In V. Uskov (Ed.), *Proceedings of* *the IASTED International Conference WEB-BASED EDUCATION* (pp. 452–457). Calgary, Canada: ACTA Press.

Huber, W., Sieprath, H., & Werth, I. (2000). Sprachdiagnostik in der Deutschen Gebärdensprache [Language diagnosis in German Sign Language]. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000 – Theorie und Anwendung* [Sign language linguistics 2000 – Theory and application] (pp. 255–276). Hamburg, Germany: Signum Verlag.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, UK: Penguin Books.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*, 523–549.

Jansma, S., Knoors, H., & Baker, A. (1997). Sign language assessment: A Dutch project. *Deafness and Education: Journal of the British Association of the Teachers of the Deaf*, *21*(3), 39–46.

Janssen, J., & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows* (6. Auflage) [Statistical data analyses with SPSS for Windows (6th ed.)] Heidelberg, Germany: Springer Verlag.

Johnston, J. (2006). *Thinking about child language – Research to practice.* Greenville, SC: Thinking Publications University – A Division of Super Duper Inc.

Johnston, J. (2007). Assessment of language learning in English speaking children. *Encyclopedia of language and literacy development* (pp. 1–9). London, Ontario, Canada: Canadian Language and Literacy Research Network. Retrieved on May 15, 2008 from http://www.literacyencyclopedia.ca/pdfs/topic.php?topId=39.

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian, and Turkish. *Journal of Child Language*, *6*, 529–545.

Johnston, T. (2004). The assessment and achievement of proficiency in a native sign language within a sign bilingual program: The pilot Auslan receptive skills test. *Deafness and Education International*, *6*(2), 57–81.

Kantor, R. (1980) The acquisition of classifiers in American Sign Language. *Sign Language Studies, 28*, 193–208.

Kiess, H. O. (1996). *Statistical concepts for the behavioral sciences* (2nd ed.). Boston, MA: Allyn & Bacon.

Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.

König, S., Konrad, R., & Langer, G. (in press). Lexikon: Der Wortschatz der DGS [The lexicon of DGS]. In H. Eichmann, M. Hansen, & J. Heßmann (Eds.), *Handbuch Deutsche Gebärdensprache – Sprachwissenschaftliche und anwendungsbezogene Perspektiven* [Handbook German Sign Language – Linguistic and applied perspectives]. Hamburg, Germany: Signum Verlag.

Krausneker, V. (2004). *Viele Blumen schreibt man "Blümer" – Soziolinguistische Aspekte des bilingualen Wiener Grundschul-Modells mit Österrreichischer Gebärdensprache und Deutsch* [Report from the Vienna bilingual elementary school model in Austrian Sign Language and German]. Hamburg, Germany: Signum Verlag.

Lane, H., Hoffmeister, R., Bahan, B. (1996). *A journey into the DEAF-WORLD.* San Diego, CA: Dawn Sign Press.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: A systematic review of the literature. *Health Technology Assessment*, *2*(9).

Leuninger, H., & Happ, D. (1997). Lena: Gebärdenspracherwerb trotz Input [Lena: Sign language development despite input]. *Frankfurter Linguistische Forschung*, *20*, 83–96.

Lewis, W. (1995). *Bilingual teaching of deaf children in Denmark – Description of a project 1982–1992*. Aalborg, Denmark: Døveskolernes Materialcenter.

Liben, L. (2006). Education for spatial thinking. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology – Vol. 4: Child psychology in practice* (6th ed.) (pp. 197–247). Hoboken, NJ: John Wiley & Sons, Inc.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lienert, G., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage) [Test construction and test analysis (6th ed.)]. Weinheim, Germany: Beltz Psychologie Verlags Union.

McNamara, T. (2000). *Oxford introductions to language study – Language testing*. Oxford: Oxford University Press.

Mahshie, S. N. (1995). *Educating deaf children bilingually – With insights and applications from Sweden and Denmark*. Washington, DC: Gallaudet University Press.

Maital, S. L., Dromi, E., Sagi, A., & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, *27*, 43–67.

Maller, S. J., Singleton, J. L., Supalla, S. J., & Wix, T. (1999). The development and psychometric properties of the American Sign Language Proficiency Assessment (ASL-PA). *Journal of Deaf Studies and Deaf Education*, *4*(4), 249–269.

Mann, W. (2008). *Facing the Challenge of Appropriately Assessing Deaf Childrens` Language Skills: An Investigation into German Deaf Childrens` Understanding of Reference in German Sign Language and in Written German*. Saarbrücken, Germany: vdm Verlag.

Mann, W., Marshall, C., Mason, K., & Morgan, G. (2010). The acquisition of sign language: The interplay between phonology and phonetics. *Language and Learning Development*, *6*, 60–86.

Mann, W., & Prinz, P. (2006). The perception of sign language assessment by professionals in deaf education. *American Annals of the Deaf*, *151*(3), 356–370.

Marschark, M. (2002). Foundations of communication and the emergence of language in deaf children. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition – Trends in language acquisition research* (pp. 1–28). Amsterdam: John Benjamins Publishing Company.

Marschark, M., Lang, H. G., & Albertini, J. A. (2002). *Educating deaf students – From research to practice.* Oxford: Oxford University Press.

Marschark, M., Schick, B., & Spencer, P. E. (2006). Understanding sign language development in deaf children. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 3–19). Oxford: Oxford University Press.

Marschark, M., & Spencer, P. E. (2006). Spoken language development of deaf and hard-of-hearing children: Historical and theoretical perspectives. In P. E. Spencer & M. Marschark (Eds.), *Advances in the spoken language development of deaf and hard-of-hearing children* (pp. 3–21). Oxford: Oxford University Press.

Martin, A. J., & Sera, M. D. (2006). The acquisition of spatial constructions in American Sign Language and English. *Journal of Deaf Studies and Deaf Education, 11*(4), 391–402.

Masataka, N. (2000). The role of modality and input in the earliest stage of language acquisition: Studies of Japanese Sign Language. In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 3–24). Mahwah, NJ: Lawrence Erlbaum Associates.

Mathur, G., & Rathmann, C. (2001). Why not GIVE-US: An articulatory constraint in signed languages. In V. Dively, M. Metzger, S. Taub, & A. M. Baer (Eds.), *Signed languages – Discoveries from international research* (pp. 1–26). Washington, DC: Gallaudet University Press.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research, 36*, 1258–1270.

Mayberry, R. I. (1995). Mental phonology and language comprehension or What does this sign mistake mean? In K. Emmorey & J. Reilly (Eds.), *Language, gesture, and space* (pp. 355–370). Mahwah, NJ: Lawrence Erlbaum Associates.

Mayberry, R. I., & Eichen, I. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language, 30*, 486–512.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, *87*, 369–384.

Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Development: Linguistic ability and early language exposure. *Nature, 417*, 38.

McCartney, E. (1993). Assessment of expressive language. In J. R. Beech, L. Harding, & D. Hilton-Jones (Eds.), *Assessment in speech language therapy* (pp. 35–48). London: Routledge.

Meier, R. P. (1987). Elicited imitation of verb agreement in American Sign Language: Iconically or morphologically determined? *Journal of Memory and Language*, *26*, 362–376.

Meier, R. P. (2002). The acquisition of verb agreement – Pointing out arguments for the linguistic status of agreement in signed languages. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 115–141). Amsterdam: John Benjamins Publishing Company.

Menyuk, P. (1988). *Language development – Knowledge and use.* Glenview, IL: Scott, Foresman, & Company.

Menyuk, P., & Brisk, M. E. (2005). *Language development and education – Children with varying language experience.* New York, NY: Palgrave Macmillian.

Miller, C., Hooper, S., & Rose, S. (2005). Avenue ASL: Developing an environment for assessing American Sign Language performance. *Advanced Technology for Learning*, *2*(3), 140–147.

Mitchell, R., & Karchmer, M. (2004): Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies* *4*(2), 138–163.

Montoya, L. A., Egnatovitch, R., Eckhardt, E., Goldstein, M., Goldstein, R. A., & Steinberg, A. G. (2004). Translation challenges and strategies: The ASL translation of a computer-based, psychiatric diagnostic interview. *Sign Language Studies*, *4*(4), 314–344.

Morgan, G. (1998). *The development of discourse cohesion in British Sign Language.* Unpublished doctoral dissertation, University of Bristol.

Morgan, G. (2000). Discourse cohesion in sign and speech. *International Journal of Bilingualism*, *4*(3), 279–300.

Morgan, G. (2002). The encoding of simultaneity in children's BSL narratives. *Journal of Sign Language and Linguistics, 5*(2), 127–161.

Morgan, G., Barrière, I., & Woll, B. (2003). First verbs in British Sign Language development. *Working Papers in Language and Communication Science: Vol. 2* (pp. 56–65). City University London.

Morgan, G., Barrière, I., & Woll, B. (2006). The influence of typology and modality on the acquisition of verb agreement morphology in British Sign Language. *First Language, 26*(1), 19–43.

Morgan, G., Herman, R., Barrière, I., & Woll, B. (2008). The onset and mastery of spatial language in children acquiring British Sign Language. *Cognitive Development*, *23*, 1–19.

Morgan, G., Herman, R., & Woll, B. (2002). The development of complex verb constructions in British Sign Language. *Journal of Child Language*, *29*, 655–675.

Morgan, G., & Woll, B. (Eds.). (2002a). *Directions in sign language acquisition – Trends in language acquisition research.* Amsterdam: John Benjamins Publishing Company.

Morgan, G., & Woll, B. (2002b). The development of complex sentences in British Sign Language. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 255–276). Amsterdam: John Benjamins Publishing Company.

Morgan, G., & Woll, B. (2003). The development of reference switch encoded through body classifiers in British Sign Language. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 297–310). Mahwah, NJ: Lawrence Erlbaum Associates.

Mounty, J. (1993). *Signed Language Development Checklist – Training Manual*. Princeton, NJ: Educational Testing Service.

Mounty, J. (1994). *Signed Language Development Checklist*. Princeton, NJ: Educational Testing Service.

Mueller Gathercole, V. C., Mon Thomas, E., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism, 11*(6), 687–720.

Newport, E. L. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 737–740). London: Macmillan Publishers Ltd.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: The MIT Press.

Newport, E., & Meier, R. (1985). The acquisition of American Sign Language. In D. Slobin (Ed.), *The cross-linguistic study of language acquisition* (pp. 881–938). Hillsdale, NJ: Lawrence Erlbaum Associates.

Niederberger, N. (2004). *Capacités langagières en Langue des Signes Française et en français écrit chez l'enfant sourd bilingue: quelles relations?* [Language competence in French Sign Language and written French in deaf bilingual children: What is the relationship?]. Unpublished doctoral dissertation, University of Geneva, Switzerland.

Niederberger, N. (2008). Does the knowledge of a natural sign language facilitate deaf children's learning to read and write? Insight from French Sign Language and written French data. In C. Plaza-Pust & E. Morales-López (Eds.), *Sign bilingualism – Language development, interaction, and maintenance in sign language contact situations* (pp. 29–50). Amsterdam: John Benjamins Publishing Company.

Norris, M. K., Juárez, M. J., & Perkins, M. N. (1989). Adaptation of a screening test for bilingual and bidialectal populations. *Language, Speech, and Hearing Services in Schools, 20*, 381–390.

Nover, S. (2005). *Working together as language planners and educational leaders: Language planning and considerations for ASL/English bilingual education*. Paper presented at the 1st Biennal ASL/English Bilingualism in Deaf Education Conference, Washington, DC, March 2005, Gallaudet University.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Oakland, T., & Lane, H. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing*, *4*(3), 239–252.

Osterlind, S. J. (2001). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Boston, MA: Kluwer Academic Publishers.

Pakendorf, C., & Alant, E. (1997). Culturally valid assessment tools: Northern Sotho translation of the Peabody Picture Vocabulary Test – Revised. *The South African Journal of Communication Disorders*, *44*, 3–12.

Papaspyrou, C., von Meyenn, A., Matthaei, M., & Herrmann, B. (2008). *Grammatik der Deutschen Gebärdensprache aus der Sicht gehörloser Fachleute* [Grammar of German Sign Language from the point of view of deaf experts]. Hamburg, Germany: Signum Verlag.

Perniss, P. (2001). *Numerus und Quantifikation in der Deutschen Gebärdensprache* [Number and Quantification in German Sign Language]. Unpublished master thesis, University of Köln, Germany.

Perniss, P. (2007). Achieving spatial coherence in German Sign Language narratives: The use of classifiers and perspective. *Lingua*, *117*, 1315–1338.

Petitto, L. A. (1987). On the autonomy of language and gesture: Evidence from the acquisition of personal pronouns in American Sign Language. *Cognition*, *27*, 1–52.

Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, *251*, 1493–1496.

Pfau, R. (2001). Typologische und strukturelle Aspekte der Negation in Deutscher Gebärdensprache [Typological and structural aspects in the negation of German Sign Language]. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000 – Theorie und Anwendung* [Sign language linguistics 2000 – Theory and application] (pp. 13–33). Hamburg, Germany: Signum Verlag.

Pfau, R. (2004). *The grammar of headshake: Sentential negation in German Sign Language*. Unpublished manuscript, University of Amsterdam, The Netherlands.

Pfau, R., & Quer, J. (2002). V-to-Neg raising and negative concord in three sign languages. *Rivista di Grammatica Generativa, 27*, 73–86.

Pfau, R., & Quer, J. (2007). On the syntax of negation and modals in Catalan Sign Language and German Sign Language. In P. Perniss, R. Pfau, & M. Steinbach (Eds.), *Visible variations – Comparative studies on sign language structure (Trends in linguistics – Studies and monographs)* (pp. 129–161). Berlin: Mounton de Gruyter.

Pfau, R., & Quer, J. (2010). Non-manuals: Their prosodic and grammatical roles. In D. Brentari (Ed.), *Sign languages* (pp. 381–402). Cambridge: Cambridge University Press.

Pfau, R., & Steinbach, M. (2005). Plural formation in German Sign Language: Constraints and strategies. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* [Sign languages: Structure, acquisition, use] (pp. 111–144). Hamburg, Germany: Helmut Buske Verlag.

Pfau, R., & Steinbach, M. (2006). Pluralization in sign and speech: A cross-modal typology study. *Linguistic Typology, 10*, 135–182.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space.* New York, NY: W. W. Norton & Company Inc.

Pizzuto, E. (2002). The development of Italian Sign Language (LIS) in deaf preschoolers. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 77–114). Amsterdam: John Benjamins Publishing Company.

Plaza-Pust, C., & Morales-López, E. (2008). Sign bilingualism: Language development, interaction, and maintenance in sign language contact situations. In C. Plaza-Pust & E. Morales-López (Eds.), *Sign bilingualism – Language development, interaction, and maintenance in sign language contact situations* (pp. 333–380). Amsterdam: John Benjamins Publishing Company.

Prinz, P. (2002). Cross-linguistic perspectives on sign language and literacy development. In R. Schulmeister & H. Reinitzer (Eds.), *Progress in sign language research – In honor of Siegmund Prillwitz* (pp. 221–234). Hamburg, Germany: Signum Verlag.

Prinz, P., Niederberger, N., Gargani, J., & Mann, W. (2005). *Cross-linguistic and cross-cultural issues in the development of tests of sign language*. Paper presented at the 10th International Congress for the Study of Child Language (ICSL), Berlin, Germany, July 25–29, 2005.

Prinz, P., Strong, M., & Kuntze, M. (1994). *The test of ASL*. Unpublished test. San Francisco: San Francisco State University, California Research Institute.

Rathmann, C. (2003). The optionality of agreement phrase: Evidence from German Sign Language (DGS). In W. E. Griffin (Ed.), *The role of agreement in natural language: Proceedings of the Fifth Annual Texas Linguistics Society Conference* (pp. 181–192). Austin, TX: Texas Linguistic Forum.

Rathmann, C., & Mathur, G. (2002). Is verb agreement the same crossmodally? In R. P. Meier, K. Cormier, & D. Quinto-Pozos (Eds.), *Modality and structure in signed and spoken languages* (pp. 370–404). Cambridge: Cambridge University Press.

Reilly, J. (2006). How faces come to serve grammar: The development of nonmanual morphology in American Sign Language. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 262–290). Cambridge: Oxford University Press.

Reilly, J., & Anderson, D. (2002). FACES: The acquisition of non-manual morphology in ASL. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 159–182). Amsterdam: John Benjamins Publishing Company.

Restrepo, M. A., & Silverman, S. W. (2001). Validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology*, *10*, 382–393.

Rosenbluth, A. R. (1976). The feasibility of test translation between unrelated languages English to Navajo. *TESOL Quarterly*, *10*(1), 33–43.

Rust, J., & Golombok, S. (2000). *Modern psychometrics – The science of psychological assessment* (2$^{nd}$ ed.). London: Routledge.

Sandler, W., & Lillo–Martin, D. (2006). *Sign language and linguistic universals*. Cambridge: Cambridge University Press.

Schembri, A. (2001). *Issues in the analysis of polycompotential verbs in Australian Sign Language (Auslan).* Unpublished doctoral dissertation, University of Sydney, Australia.

Schembri, A. (2003). Rethinking "classifiers" in signed languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 3–34). Mahwah, NJ: Lawrence Erlbaum Associates.

Schembri, A., Wigglesworth, G., Johnston, T., Leigh, G., Adam, R., & Baker, R. (2002). Issues in development of the Test Battery for Australian Sign Language Morphology and Syntax. *Journal of Deaf Studies and Deaf Education*, *7*(1), 18–40.

Schick, B. (1987). *The acquisition of classifier predicates in American Sign Language.* Unpublished doctoral dissertation, Purdue University, West Lafayette, Indiana (ProQuest UMI number: 8807670).

Schick, B. (1990). The effects of morphosyntactic structure on the acquisition of classifier predicates in ASL. In C. Lucas (Ed.), *Sign language research – Theoretical issues* (pp. 358–374). Washington, DC: Gallaudet University Press.

Simon, A. J., & Joiner, L. M. (1976). A Mexican version of the Peabody Vocabulary Test. *Journal of Educational Measurement*, *13*(2), 137–143.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, *49*, 370–407.

Singleton, J. L., & Supalla, S. (2003). Assessing children's proficiency in natural signed languages. In M. Marschark & P. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 289–302). Oxford: Oxford University Press.

Solarsh, B., & Alant, E. (2006). The challenge of cross-cultural assessment – The test of ability to explain for Zulu-speaking children. *Journal of Communication Disorders*, *39*, 109–138.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York, NY: Holt, Rinehart, & Winston.

Slobin, D. I., Hoiting, N., Kuntze, M., Lindert, R., Weinberg, A., Pyers, J., et al. (2003). A cognitive/functional perspective on the acquisition of "classifiers". In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 271–296). Mahwah, NJ: Lawrence Erlbaum Associates.

Snitzler Reilly, J., McIntire, M., & Bellugi, U. (1991). Baby face: A new perspective on universals in language acquisition. In P. Siple & S. D. Fischer (Eds.), *Theoretical issue in sign language research – Vol. 2 Psychology* (pp. 9–24). Chicago, IL: The University of Chicago Press.

Spencer, P. E., & Harris, M. (2006). Patterns and effects of language input to deaf infants and toddlers from deaf and hearing mothers. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 71–101). Oxford: Oxford University Press.

Strong, M., & Prinz, P. (1997). A study of the relationship between ASL and English literacy. *Journal of Deaf Studies and Deaf Education*, *2*(1), 37–46.

Strong, M., & Prinz, P. (2000). Is American Sign Language skill related to English literacy? In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 131–142). Mahwah, NJ: Lawrence Erlbaum Associates.

Supalla, T. (1982). *Structure and acquisition of verbs of motion and location in American Sign Language.* Unpublished doctoral dissertation, University of California, San Diego.

Supalla, T., Newport, E., Singleton, J., Supalla, S., Coulter, G., & D. Metlay (1995). *An overview of the Test Battery for American Sign Language Morphology and Syntax*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), April 20, 1995, San Francisco, CA.

Supalla, T., Newport, E., Singleton, J., Supalla, S., Coulter, G., & D. Metlay (unpublished). *The Test Battery for American Sign Language Morphology and Syntax*.

Surian, L., & Tedoldi, M. (2005). *Adaptation of BSL Receptive Skills Test to Italian Sign Language (LIS).* Unpublished manuscript, University of Trieste, Department of Psychology, Italy.

Sutton-Spence, R., & Woll, B. (1999). *The linguistics of British Sign Language – An introduction.* Cambridge: Cambridge University Press.

Tang, G., Sze, F., & Lam, S. (2007). Acquisition of simultaneous constructions by deaf children of Hong Kong Sign Language. In M. Vermeerbergen, L. Leeson, & O. Crasborn (Eds.), *Simultaneity in signed languages: Form and function* (pp. 283–316). Amsterdam: John Benjamins Publishing Company.

Thordardottir, E. T., & Ellis Weismer, S. (1996). Language assessment via parent report: Development of a screening instrument for Icelandic children. *First Language*, *16*, 265–285.

Tomasello, M. (2003). *Constructing a language – A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Van de Vijver, F., & Leung, K. (1997a). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Vol. 1* (2nd ed.) (pp. 257–300). Boston, MA: Allyn & Bacon.

Van de Vijver, F., & Leung, K. (1997b). *Methods and data analysis of cross-cultural research*. Newbury Park, CA: Sage Publication.

Van de Vijver, F., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39–64). Mahwah, NJ: Lawrence Erlbaum Associates.

Van Dyk, T., & Weideman, A. (2004). Switching construct: On the selection of an appropriate blueprint for academic literacy assessment. *Journal of Language Teaching, 38*(1), 1–13.

Van Hoek, K., O'Grady, L., & Bellugi, U. (1987). Morphological innovation in the acquisition of American Sign Language. *Papers and reports on child language development* (pp. 116–123). Palo Alto, CA: Stanford University Press.

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, *40*(4), 1079–1087.

Wiig, E. H., & Secord, W. A. (forthcoming). *Making sense of test results – A practical guide*. Arlington, TX: White Paper. Knowledge Research Institute, Inc.

Wisch, F.-H. (1990). *Lautsprache und Gebärdensprache – Die Wende zur Zweisprachigkeit in Erziehung und Bildung Gehörloser* [Spoke language and sign language – The turn to bilingual education of the Deaf]. Hamburg, Germany: Signum Verlag.

White, A., & Tischler, S. (1999). Receptive sign vocabulary tests: Tests of single-word vocabulary or iconicity. *American Annals of the Deaf*, *144*(4), 334–338.

Woll, B. (1998). Development of signed and spoken language. In S. Gregory, P. Knight, W. McCracken, S. Powers, & L. Watson (Eds.), *Issues in deaf education* (pp. 58–68). London: David Fulton Publishers.

Woolfe, T., Herman, R., Roy, P., & Woll, B. (2010). Early lexical development in native signers: A BSL adaptation of the MacArthur-Bates CDI. *Child Psychology and Psychiatry*, *51*(3), 322–331.

Zeshan, U. (2004). Hand, head, and face – Negative constructions in sign languages. *Linguistic Typology*, *8*, 1–58.

Zeshan, U. (Ed.). (2006). *Interrogative and negative constructions in sign languages* (Sign language typology series). Nijmegen, The Netherlands: Ishara Press.

# Appendices

## Appendix A-1
## Changes to BSL Test Materials (January 22, 2005)

### (1) Vocabulary check (22 items)

| *Gloss of item* | *Item number* |
|---|---|
| KISTE | 5 |
| KOPFHÖRER | 14 |
| HÖRGERÄT | 15 |
| MUTTER/FRAU | 18 |
| TISCH | 20 |

### (2) Morpho-syntactic items (numbering based on original BSL test)

| *Item number/name* | *Changes to picture(s)* |
|---|---|
| P3-TEDDY KLEIN | P3.3*: The teddy needs to be smaller |
| 13-KIND SCHAUT-AUF/HOCH | The color of the dress should be the same in all four pictures |
| 20-JUNGE-SCHLA-GEN MÄD-CHEN-WIRD-GESCH-LAGEN | 20.4: Cartoon-like “lines” are missing expressing pain |
| 24-SCHLANGE-LEUTE | 24.1: Replace the words “Bus Stop” with a picture of a bus<br>24.2: Add two more people to the queue and replace the words “Bus Stop” with a picture of a bus |
| 25-REGENSCHIRM-OFFEN-HALTEN GEHEN | 25.1: Change the legs of the woman to make it clearer that she is standing (not walking) |
| 29-MUTTER BRIEF GEBEN | 29.1: Replace the mailbox with a German yellow and square mailbox |

| *Gloss of item* | *Item number* |
|---|---|
| 32-KIND BUCH-ZEIGEN-ZUR-SEITE | 32.2-32.4: Improve the action of "showing-to-side" better in these three pictures |
| 35-BECHER/TASSE HERUNTERFALLEN-NICHT | 35.1: Make it clearer that something is falling<br>35.4: The color of the glass should have the same color as in the other pictures |
| 36-HÖRGERÄT OHNE | 36.1: Add ball<br>36.2: Take ball away<br>36.4: Replace with ball (take picture from vocabulary check) |
| 38-REIHE-AUTO UNTEN (LINKS) | 38.1: A point of reference is missing: for example, a sign in front to the right and/or lines indicating the parking space<br>38.2: Same as above |
| 39-HUND-LIEGEN-INNEN-RECHTS (KISTE) | 39.1-39.4: The box should be the same size in all four pictures; make the dog smaller (same size in all four pictures). |

* The pictures (potential answers) of the items are numbered clockwise: (1) upper left, (2) upper right, (3) lower left, and (4) lower right. For example, the lower left picture for practice Item 3 is numbered as P3.3.

# Appendix B-1
# Questionnaire for Pilot 1 (Translation)[43]

## Questionnaire – Personal data

### I. Background information
1. First name and family name?
2. How old are you (age)?
3. Where did you grow up? (Grew up in …)
4. Where do you live now?
5. Where else have you lived?
6. Are your parents Deaf and/or hearing?
7. Are your brothers and sisters Deaf and/or hearing?
8. Do you have other Deaf relatives, for example, aunts and uncles?
9. Is your partner/spouse Deaf or hearing?
10. Are your children Deaf and/or hearing?

### II. Educational background
11. Did you attend a kindergarten for the Deaf?
12. Did you attend a school for the Deaf?
13. What kind of apprenticeship did you complete and where?
14. What is your current job/position?

### III. Communication in sign language (in the past)
15. When did you start to use sign language? For example, at kindergarten or even earlier? Or at school?
16. Where did you learn sign language?
17. From whom did you learn sign language?
18. How did you communicate at home in the family? (Check which follow-up question fits, depending if parents are hearing or Deaf.)
    a) How did you and your hearing parents communicate? OR
    b) How did your Deaf parents and your hearing siblings communicate?
    c) How did your Deaf parents and your Deaf siblings communicate?
19. Do your parents and/or siblings sign? Can you judge/estimate if they can sign well or not so well?

[43] Only the translated version (in English) of the original German questionnaires (and other materials) are available as appendices.

### IV. Communication in sign language (today)

20. How do you communicate today in your own family?
    a) How do you communicate with your partner/spouse?
    b) How do you communicate with your Deaf and/or hearing children?
21. Do you have Deaf colleagues at work?
22. Generally speaking, how do you communicate with hearing people? Pointing, writing, lip-reading…?
23. Which means/forms of communication do you use, for example, DGS, LBG?
24. On an everyday basis, when and where do you use sign language? For example in the family, with friends, Deaf club…?
25. Do you use a sign language interpreter, in a situation such as a doctor's appointment, or do hearing friends interpret for you?

## Appendix C-1
## Consent Form for Deaf Adults for Pilot 1 (Translation)

### Consent form

Hereby, I agree [name of person] that the video-recordings of me from [date] can be used by Tobias Haug for his dissertation project.

I also agree that the video recordings can be shown at conferences or seminars.

[Signature and date]

# Appendix C-2

## Consent Form for Children for Pilot 1 (Translation)

### Consent for Pilot 1 of the DGS Receptive Skills Test

I (parent/legal guardian), [name of parent], agree that the video recordings of my child [name of child] from September 7, 2004 …. (please check the appropriate box).

1. ❑ can be used for the dissertation of Tobias Haug.
2. The data (i.e., the video recordings), but not the personal data obtained from the interview/questionnaire, can be used at conferences or seminars.
   ❑ yes
   ❑ no, the video recordings should be treated as confidential.
3. My personal data (name, DOB) should be treated as confidential.
   ❑ yes
   ❑ no
4. I would like to be acknowledged as an informant in the dissertation.
   ❑ yes
   ❑ no

[Date and signature of parents/legal guardian]

# Appendix D-1
# Regional Variations and Conventionalized/ Unconventionalized Forms of Vocabulary Items for Pilot 1 (*N* = 13)

(Form variants are numbered as they occur in the DGS-VT data bank)

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| APFEL (apple) | 1 | APFEL1<br>APFEL3 | 13/13<br>1/13 | APFEL2: Depicts eating of unspecified object. | 1/13 |
| BALL (ball) | 2 | BALL1 | 13/13 | | |
| BETT (bed) | 3 | BETT3, 5, 6, 7: BETT3, 5, and 6 refer to the same sign, only produced (1) with one hand only (BETT5) instead of two hands (BETT3), and (2) both hand were located on the same level and below each other (BETT6), respectively. | 5/13, 2/13, 1/13, 1/13 | BETT1, 2, 4: Eliciting the verb sleeping. | 1/13, 1/13, 3/13 |
| BUCH (book) | 4 | BUCH2, 3: BUCH3 with mouthing *Heft* (booklet). | 12/13, 2/13 | BUCH1: Eliciting the action of opening a book. | 1/13 |
| KISTE (box) | 5 | KISTE1, 2, 3, 4, 5: Each revealing different aspects of the box (whereas only variation 1-3 can be considered | 2/13, 8/13, 1/13, 1/13, 1/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| | | referring to the actual object box). Variation 4 and 5 referred to a parcel or a flat box. Variation 1, 2, or 3 will be used for the test. The different variant also produced different mouthing (i.e., *Schachtel*, *Karton*, not *Kiste*). | | | |
| JUNGE (boy) | 6* | JUNGE1: Produced by the majority of the informants in the South-West of Germany, in 2 cases with Southern German mouthing pattern (*Bub* instead of *Junge*). It could be considered as a dialect variant. | 6/13 | JUNG6: One informant produced a different manual sign JUNG (*young*) instead of the sign JUNGE | 1/13 |
| | | (JUNGE2: Considered a phonological alteration of the variant JUNGE4). | 1/13 | | |
| | | JUNGE3: Most of the informants (6/7) were from the South-West and South, three of these informants accompanied the sign with a Southern German | 7/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| | | mouthing pattern (*Bub* instead of *Junge*). | | | |
| | | JUNGE4: Produced by informants from Northern Germany and could be clearly analyzed as a regional Northern German sign. | 3/13 | | |
| | | JUNGE5: Produced by informants from the South with Southern German mouthing pattern (*Bub*). This sign can be considered as a dialect variation for the South. | 2/13 | | |
| KIND (child) | 7* | KIND1, 2: Neither variant could be assigned to a specific region (South and South-West). The stimuli elicited JUNGE more often than KIND. | 4/13, 2/13 | | |
| | | Only variant KIND1 will be used in the actual test. | | | |
| AUTO (car) | 8 | AUTO1 | 13/13 | | |
| JACKE / MANTEL (jacket) | 9 | MANTEL2 | 12/13 | MANTEL1 | 1/13 |
| | | MANTEL3: Produced by only one | 1/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| | | informant, using the same sign as variant MANTEL2 but a different mouthing pattern (*Jacke*). | | | |
| TASSE (cup) | 10 | TASSE1 | 9/13 | TASSE2: Eliciting the verb for drinking | 3/13 |
| | | | | TASSE3: Eliciting a handshape for a cup standing somewhere, not the lexical sign. | 1/13 |
| HUND (dog) | 11** | HUND1: Used by informants from the South-West and Northern part of Germany, the majority (6/8) from South-West. | 8/13 | HUND3, 5, 6: not acceptable forms. | 1/13, 1/13, 1/13 |
| | | HUND2: Could not be assigned to a specific region. | 3/13 | | |
| | | HUND4: Could be assigned to a specific region (South). | 3/13 | | |
| HALSBAND (collar) | 12** | HALSBAND1, 2, 3: Three variants, with minor differences (one hand vs. two-hand sign) | 4/13, 3/13, 1/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| MÜTZE (hat) | 13 | MÜTZE1 | 10/13 | | |
| | | MÜTZE2: Conventionalized form but produced one-handed. | 1/13 | | |
| KOPFHÖRER (headphone) | 14 | KOPFHÖRER1, 3: Could not be assigned to a specific region. | 7/13, 5/13 | KOPFHÖRER2: Describing the event of listening to music. | 1/13 |
| HÖRGERÄT (hearing aid) | 15 | HÖRGERÄT1, 2: Conventionalized forms with phonological variations. | 8/13, 3/13 | HÖRGERÄT3 | 1/13 |
| EIS (ice cream) | 16 | EIS1 | 13/13 | | |
| BRIEF (letter) | 17 | BRIEF1 (3 informants did not produce a sign). | 10/13 | | |
| MUTTER / FRAU (mother / woman) | 18 | MÄDCHEN2 (girl): Informant from the South-West | 1/13 | MÄDCHEN1 | 1/13 |
| | | | | MAMA1 | 1/13 |
| | | MÄDCHEN3: Informants from the North | 3/13 | | |
| | | FRAU1: Form could not be assigned to a specific region (South and South-West). | 5/13 | | |
| | | FRAU2: Form could not be assigned to a specific region (South and South-West). | 2/13 | | |
| | | FRAU3: Form could not be assigned to | 4/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| | | a specific region (North, South, South-West). | | | |
| | | FRAU4: Informant from the North. | 1/13 | | |
| | | FRAU5: Informants from South, manual sign produced with different mouthing pattern (*Dame*) (lady) | 2/13 | | |
| | | MUTTER1: Informants from the South. | 3/13 | | |
| | | (This stimulus produced additionally to the sign for MAMA and FRAU also the sign for MÄDCHEN (girl)) | | | |
| BLEISTIFT (pencil) | 19 | BLESISTIFT1: Produced by informants from all regions. | 11/13 | | |
| | | BLEISTIFT2: Produced by informants from the South. | 2/13 | | |
| TISCH (table) | 20 | TISCH2 | 10/13 | TISCH1: Form a description rather than a conventionalized form. | 3/13 |
| TEDDYBÄR (teddy) | 21 | TEDDYBÄR3, 4, 5: Using the compound sign TEDDY+BÄR. | 2/13, 1/13, 4/13 | | |

| *Vocabulary item* | *Item number* | *Conventionalized forms* | *Produced by number of informants (N = 13)* | *Unconventionalized forms* | *Produced by number of informants (N = 13)* |
|---|---|---|---|---|---|
| | | TEDDYBÄR1, 2, 6: The sign BÄR (bear) was elicited, not the sign for teddy, but each with a slight phonological alteration (movement). | 4/13, 1/13, 1/13 | | |
| REGENSCHIRM (umbrella) | 22 | REGENSCHIRM1 | 11/13 | | |
| | | REGENSCHIRM2: Using the compound sign REGEN+SCHIRM. | 2/13 | | |

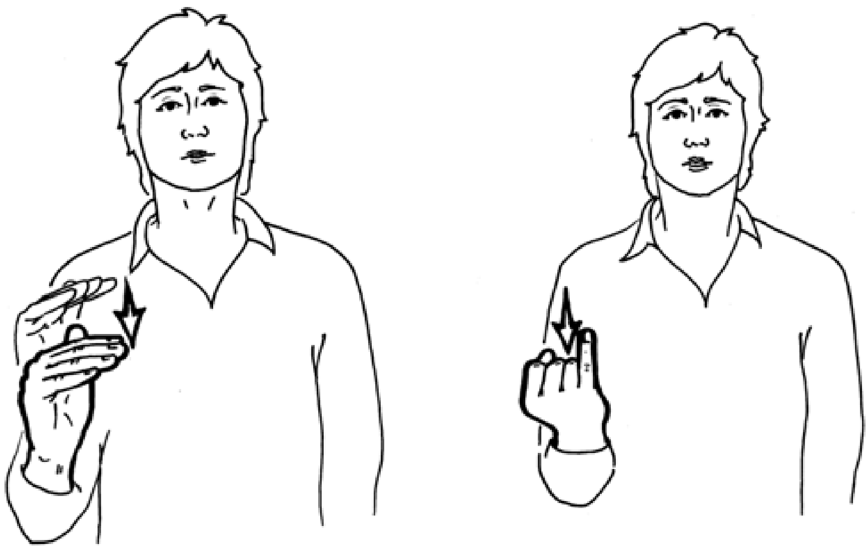
* The same picture was used for the sign JUNGE (boy) and KIND (child).

** The same picture was used for the sign HUND (dog) und HALSBAND (collar).

# Appendix D-2
# Examples of Regional Variations in Pilot 1

Example 1: HUND1 (dog) (top left), HUND2 (top right), HUND4 (below)

Example 2: KIND1 (child) (left) and KIND2 (right)



# Appendix D-3
# Complete List of Items for the DGS Receptive Skills Test (First Version)

## Items for DGS Receptive Skills Test

(Items 1-40 are based on BSL Receptive Skills Test)
(List from Herman, 2002)

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| | | *Practice items:* | | |
| P1 | Two sign combination | CHILD EAT<br>child eating plate of food | 1 **2**<br>3 | child with plate of food, not eating;<br>child drinking |
| P2 | Two sign combination | MUMMY READ<br>woman sitting down reading | **1**<br>2 3 | child writing;<br>child sitting down reading |
| P3 | Size & shape-specifier | TEDDY SMALL<br>small teddy | 1 2<br>**3** 4 | big teddy;<br>small pencil;<br>small cup |
| | | *Main test items:* | | |
| 1 | Number/ distribution | LOTS APPLE<br>lots of apples | 1 **2**<br>3 4 | a few apples;<br>one apple;<br>person carrying heavy shopping |
| 2 | Number/ distribution & spatial verb morphology | CAR ROW ROW ROW<br>three rows of parked cars | 1 2<br>**3** 4 | row of four parked cars;<br>shelves of books;<br>one parked car |
| 3 | Negation | ICE-CREAM NOTHING<br>child with no ice-cream (w/hat) | 1<br>2 **3** | child with ice cream but no hat;<br>child with ice-cream and hat |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| 4 | Negation | NOT-LIKE EAT<br>child rejecting food | **1** 2<br>3 4 | child rejecting drink;<br>child with plate of food, not eating;<br>child enjoying eating food |
| 5 | Spatial verb morphology | BOOK ON<br>book on bed | 1 **2**<br>3 | brush on bed;<br>book under bed |
| 6 | Number/<br>distribution | ONE-TEDDY<br>one teddy | **1** 2<br>3 | rows of teddies;<br>group of three teddies |
| 7 | Noun-verb agreement | DRIVE (VERB)<br>person driving a car | 1 **2**<br>3 4 | empty parked car;<br>person driving a train;<br>person riding a bike |
| 8 | Negation | HAT NOTHING<br>snowman with no hat | 1 2<br>**3** 4 | child with hat but no shoes;<br>snowman with hat;<br>child with hat and shoes |
| 9 | Spatial verb morphology | BALL TABLE ON<br>ball on table | **1** 2<br>3 4 | doll on table;<br>ball under table;<br>ball on chair |
| 10 | Spatial verb morphology<br>(spatial verb & movement classifier) | TWO-PEOPLE-MEET<br>man and woman walking towards each other | 1 2<br>3 **4** | man and woman standing beside each other;<br>man and woman walking away from each other;<br>man follows woman |
| 11 | Spatial verb morphology | DOG IN<br>dog's head visible from the open top of a box | 1<br>2 **3** | dog under a box;<br>dog on top of a box |
| 12 | Spatial verb morphology & Number/ distribution | PERSON-GO-DOWN-ESCALATOR<br>man standing on descending escalator | 1 **2**<br>3 4 | man standing on ascending escalator;<br>group of people descending escalator; |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| | | | | two people standing on descending escalator |
| 13 | Spatial verb morphology<br>(agreement verb) | CHILD LOOK-UP<br>boy seated on the floor looks up at his mother seated on a chair | 1 2<br>3 **4** | mother looks down at child playing;<br>child looks at mother while she is reading;<br>mother is sitting on the floor and looking at the standing boy |
| 14 | Number/ distribution | FEW-CUPS<br>three cups | 1 2<br>**3** | one cup;<br>rows of cups |
| 15 | Spatial verb morphology | CAR BEHIND<br>house with car parked behind | **1** 2<br>3 4 | car parked in front of house;<br>car parked alongside house/left of the house;<br>car on its own |
| 16 | Size & shape-specifier | CURLY-HAIR<br>person with long curly ringlets | 1 2<br>**3** 4 | long wavy hair;<br>short straight hair;<br>long frizzy hair |
| 17 | Spatial verb morphology | BOX UNDER BED<br>box visible under bed | 1 **2**<br>3 4 | box under table;<br>teddy under bed;<br>box on bed |
| 18 | Spatial verb morphology<br>(agreement verb) | BOOK GIVE-TO-CHILD<br>mother gives book to child | 1 2<br>3 **4** | mother holds book;<br>child gives book to mother;<br>child holds book |
| 19 | Noun-verb agreement | BOY-DRINK (VERB)<br>boy drinking from cup | **1** 2<br>3 4 | cup;<br>can of coke;<br>person talking on phone |
| 20 | Spatial verb morphology (spatial verb + body classifier) | BOY-HIT GIRL-GET-HIT<br>boy punches girl's face | **1** 2<br>3 4 | boy and girl "crash" with their heads;<br>girl with bruised face;<br>girl punches boy's face |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| 21 | Size & shape - specifier | PENCIL THICK<br>thick pencil | 1 **2**<br>3 4 | thin book;<br>thick book;<br>thin pencil |
| 22 | Size & shape-specifiers | THICK-STRIPES-DOWN-TROUSERS<br>trousers with thick vertical stripes | **1** 2<br>3 4 | trousers with horizontal stripes;<br>trousers with thin horizontal stripes;<br>trousers with thin vertical stripes |
| 23 | Negation | NOT-SLEEPING<br>child reading in bed | **1**<br>2 3 | child sleeping in bed;<br>baby sleeping in cot |
| 24 | Number/<br>distribution | QUEUE<br>queue of people at bus stop | 1 **2**<br>3 4 | single person standing at bus stop;<br>crowd of people;<br>fence |
| 25 | Handling classifier | HOLD-UMBRELLA-OPEN-WALK<br>woman holding open umbrella walking away from a house | 1 2<br>3 **4** | woman holding open umbrella, standing;<br>woman holding closed umbrella, walking;<br>woman walking, umbrella lying open in front of the door |
| 26 | Noun-verb agreement | PENCIL<br>pencil | 1 2<br>**3** 4 | child writing with pencil;<br>open book;<br>person painting picture |
| 27 | Spatial verb morphology<br>(spatial verb + body part classifier) | (BOY-left) POUR-WATER-OUT (BOY-right) WATER-POUR HAIR-WET<br>two boys in bath, one pours water on other's hair | 1 2<br>3 **4** | boy pours water on his hair;<br>boy pours water on floor by bath;<br>two boys in bath, one pours water on the other boy |
| 28 | Negation | HEADPHONE NOTHING<br>child with no headphones, but with drink | 1<br>**2** 3 | child wearing headphones and drink;<br>child with headphones and no drink |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| 29 | Spatial verb morphology<br>(agreement verb) | MUMMY LETTER GIVE<br>woman hands letter to man standing at her side | 1 2<br>3 **4** | woman posts letter;<br>woman holding letter, standing closed to a door;<br>man hands letter to a woman standing to her side |
| 30 | Negation<br>(embedded clause) | CHILD COAT RAIN NOTHING<br>child wearing a coat in the sunshine | 1 **2**<br>3 4 | child wearing a coat in the rain;<br>child with no coat in the sunshine;<br>child with coat in the rain |
| 31 | Negation | CAN'T-REACH<br>small child reaching up for teddy on top of cupboard | 1 **2**<br>3 | taller child taking teddy from top of cupboard;<br>small child climbing chair |
| 32 | Spatial verb morphology<br>(agreement verb) | CHILD BOOK SHOW-TO-SIDE<br>boy and girl sitting on the floor next to each other, boy shows girl book | 1 2<br>3 **4** | mother reads book;<br>mother shows child book;<br>child shows mother book |
| 33 | Negation<br>(embedded clause) | DOG NO COLLAR EAT-BIG-BONE<br>small dog with no collar eating big bone | **1** 2<br>3 4 | small dog with collar eating a big bone;<br>small dog with collar eating small bone;<br>small dog with no collar eating small bone |
| 34 | Spatial verb morphology<br>(spatial verb) | DOG-IN-FRONT<br>dog lying in front of a box | 1 2<br>**3** 4 | dog sitting next to box;<br>dog behind box;<br>dog walking away from box |
| 35 | Negation & handling classifier | NOT-DROP-CUP<br>child holding cup carefully | 1 **2**<br>3 4 | child with broken cup on floor;<br>child holding a cup towards a boy;<br>broken cup |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| 36 | Negation | BALL NOTHING<br>dog with no ball | 1 **2**<br>3 4 | child/boy with (hearing aid) and ball;<br>child/girl with (hearing aid) and ball;<br>ball |
| 37 | Handling classifier<br>(spatial verb + handling classifier) | EAT-THIN-SANDWICH<br>boy eating a thin sandwich | 1 **2**<br>3 4 | boy eating biscuit;<br>boy eating crisps;<br>boy eating large burger/thick sandwich |
| 38 | Spatial verb morphology<br>(spatial verb) | ROW-OF-CARS BOTTOM-LEFT<br>row of parked cars at the bottom of the picture | **1** 2<br>3 4 | row of cars at the top of the picture;<br>two cars behind each other;<br>single car (on road) |
| 39 | Spatial verb morphology<br>(spatial verb) | DOG-LIE-INSIDE-LEFT<br>dog lying inside box to the left | **1** 2<br>3 4 | dog on top of box on left;<br>dog on top of box on right;<br>dog inside box to the left |
| 40 | Spatial verb morphology<br>(spatial verb) | HOUSE TOP-RIGHT<br>crossroads with a house in the top right quadrant | 1 2<br>3 **4** | house in the top left quadrant;<br>house in bottom left quadrant;<br>house in bottom right quadrant |
| 41 | Number/ distribution | CARS LOTS<br>many cars/rows of cars | **1** 2<br>3 4 | two cars;<br>one car;<br>one row of cars (4 cars) |
| 42 | Number/ distribution | PENCILS FEW<br>a few pencils (3) | 1 2<br>**3** 4 | one pencil;<br>lots of pencils (i.e., "rows");<br>thick pencil |
| 43 | Noun/verb agreement | CHILD SIT<br>child sitting on a chair | 1 2<br>3 **4** | woman sitting on a chair;<br>chair;<br>child standing |

| | *Linguistic feature assessed* | *Gloss of DGS receptive target and description of picture* | *R* | *Descriptions of distractor pictures (lexically, syntactically or phonologically contrastive)* |
|---|---|---|---|---|
| 44 | Spatial verb morphology | TEDDY UNDER (BED)<br>teddy is (visible) under the bed | 1 **2**<br>3 4 | book under bed;<br>teddy under table;<br>teddy on bed |
| 45 | Spatial verb morphology | DOG BEHIND (BOX)<br>dog is behind a box, the head is visible | 1 2<br>**3** 4 | dog in front of box;<br>dog on top of box;<br>dog inside box |
| 46 | Size/shape-specifiers | CHILD SWEATER STRAIGHT-ROW-DOTS<br>child with a sweater with colored dots in rows | **1** 2<br>3 4 | child with sweater with thin stripes;<br>child with sweater with colored dots, in random order;<br>child with a sweater with colored big dots in rows |
| 47 | Number/ distribution | ONE BALL<br>one ball | 1 **2**<br>3 | rows of balls (lots of balls);<br>few balls/three balls |
| 48 | Spatial verb morphology | CHILD STANDING-FRONT-OF-CAR<br>child is standing in front of a car | **1** 2<br>3 4 | child standing behind the car;<br>child standing beside the car;<br>child stands |
| 49 | Size/shape-specifiers | PENCIL SMALL<br>small pencil | 1<br>2 **3** | thick pencil;<br>regular pencil |
| 50 | Negation | BOY WITHOUT DRINK<br>boy without a drink, but ice-cream | 1 **2**<br>3 | child with drink, without hat;<br>child with drink and headphones |

Key to symbols used:

R = response options (numbers indicate number and arrangement of pictures on page in test booklet; emboldened number represents target picture)

# Appendix E-1
# Evaluation Sheet for Vocabulary Check (Translation)

(from Herman et al., 1999)

## German Sign Language Receptive Skills Test (DGS-VT)

## Evaluation Sheet for Vocabulary Check

Date:______________________________

ID of child:_________________________

Test administrator:____________________

| *Vocabulary* | | *Does the child know the sign?* *Please check off the appropriate cell.* | | | |
|---|---|---|---|---|---|
| | | ✔ | ✖ | ∅ | *Comments* |
| 1 | APFEL | | | | |
| 2 | BALL | | | | |
| 3 | BETT | | | | |
| 4 | BUCH | | | | |
| 5 | KISTE | | | | |
| 6 | JUNGE* | | | | |
| 7 | KIND* | | | | |
| 8 | AUTO | | | | |
| 9 | JACKE | | | | |
| 10 | TASSE | | | | |
| 11 | HUND+ | | | | |
| 12 | HALSBAND+ | | | | |
| 13 | MÜTZE | | | | |
| 14 | KOPFHÖRER | | | | |
| 15 | EIS | | | | |

| *Vocabulary* | | *Does the child know the sign?*<br>*Please check off the appropriate cell.* | | | |
|---|---|---|---|---|---|
| 16 | BRIEF | | | | |
| 17 | MUTTER/FRAU | | | | |
| 18 | BLEISTIFT | | | | |
| 19 | TISCH | | | | |
| 20 | TEDDY | | | | |
| 21 | REGENSCHIRM | | | | |
| *Number of signs produced correctly* | | | | | |

*The same picture was used for the sign JUNGE and KIND
+ The same picture was used for the sign HUND and HALSBAND

## Appendix F-1
## Consent Form for Deaf Adults for Pilot 2 (Translation)

### Consent form for participating in a study with the DGS Receptive Skills Test

I, [name of person] agree that (1) the information obtained by the questionnaire, and (2) the data of the sign language test will be used for the dissertation of Tobias Haug.
I have been informed that collected data (questionnaires, results of sign language test) will be treated as confidential.
I would like to be acknowledged as informant in the dissertation.
❑ yes
❑ no

[Date and signature]

# Appendix F-2
# Background Questionnaire for Deaf Adults for Pilot 2 (Translation)

Project Sign Language Tests

German Sign Language Receptive Skills Test
Questionnaire for Pilot with Deaf adults

1. Name: ______________________________

2. Age: ______________________________ years

3. Are the following members of your family Deaf, hard-of-hearing, or hearing?

| Persons in your family | Deaf (D), hard-of-hearing (hoh), or hearing (h)? |
|---|---|
| 1. Mother | |
| 2. Father | |
| 3. Brother(s) | |
| 4. Sister(s) | |
| 5. Grandmother | |
| 6. Grandfather | |
| 7. Partner/spouse | |
| 8. Own children | |
| 9. Aunts/uncles | |
| 10. Cousins | |
| 11. Other | |

4. Did you attend a kindergarten for the Deaf?
- ❑ yes
- ❑ no

5. Did you attend a school for the Deaf?
- ❑ yes
- ❑ no

6. What kind of apprenticeship or university/college program did you complete?

7. Where did you learn sign language? (multiple answers possible)
- ❑ at home
- ❑ at kindergarten
- ❑ at school
- ❑ other_______________________________________________

8. From whom did you learn sign language? (multiple answers possible)
- ❑ from my parents (and/or other members of my family)
- ❑ from my siblings
- ❑ from Deaf friends at kindergarten
- ❑ from Deaf friends at school
- ❑ other_______________________________________________

9. Do the following members of your family know sign language?

| Persons in your family | Sign language proficiency? |
|---|---|
| 1. Mother | |
| 2. Father | |
| 3. Brother(s) | |
| 4. Sister(s) | |
| 5. Grandmother | |
| 6. Grandfather | |
| 7. Partner/spouse | |
| 8. Own hearing children | |
| 9. Own Deaf children | |
| 10. Aunts/uncles | |
| 11. Cousins | |
| 12. Other | |

10. How do you communicate with hearing people?
- ❑ pointing
- ❑ writing

❑ lip-reading
❑ other________________________________________

11. Which other forms of communication do you use? (multiple answers possible)
❑ German Sign Language (DGS)
❑ Signed German (LBG)
❑ Finger spelling
❑ other________________________________________

12. In everyday life, when and where do you use sign language? (multiple answers possible)
❑ in the family
❑ with friends
❑ at Deaf club
❑ at work
❑ other________________________________________

I would like to be informed about the results of the Pilot 2 study.
❑ yes
❑ no

Thanks a lot for your collaboration!

# Appendix F-3
# Results of Pilot 2 with Deaf Adults (*N* = 5)

Comments from the five Deaf adults who took part in the Pilot 2 test and suggested revisions for next version

| *Item No** | *Item*** | *Comments by Deaf adults* | *Researcher's conclusion/ comments and resulting revisions of stimuli* |
|---|---|---|---|
| 2 | CAR-ROW-ROW-ROW | DA01: Chose bookshelf picture as movement resembles more that for a bookshelf or a table than for rows of cars<br>DA03: Same as DA01, movement of car-row-row-row is not correct, is rather confusing & subject is missing<br>DA05: Movement is not correct, refers more to 2.2 (i.e., bookshelf) | *Video*: The movement was changed, less tracking, and more "tapping" movement (i.e., a small downward movement repeated as the hand moves from left to right) |
| 11 | DOG-IN | DA03: Not really DGS, which would be first BOX CL-box then DOG CL-dog-inside<br>DA05: Word order, BOX in (with one finger) | *Video*: Wrong word order, its should be either (a) BOX, indicating with the index finger IN where the box was signed, or (b) CL-DOG-IN-BOX and a reference for the side of the box with non-dominant hand, i.e. BOX CL-BOX-left DOG CL-DOG-IN-BOX-right, alternative (b) was used |
| 15 | CAR-BEHIND | DA05: Wrong handshape referring to the house. Instead of the 5-Clawed handshape, only the flat hand should be used. Also movement of car is wrong. Right now it looks more like "driving" than of "being" | *Video*: Continue to use 5-Clawed proform for HOUSE-is-there but for CAR-is-located use a clearer movement for "is located there" because the other four Deaf adults considered the |

| *Item No** | *Item*** | *Comments by Deaf adults* | *Researcher's conclusion/ comments and resulting revisions of stimuli* |
| --- | --- | --- | --- |
| | | | handshape to be correct<br>*Pictures*: 15.1 changed so that the car is behind and not in the house |
| 30 | CHILD COAT RAIN NOTHING | DA03: Could be both 30.3 and 30.4. (Picture 30.3 depicts a girl standing (without a rain coat) and Picture 30.4 depicts a girl walking in the rain) | *Video*: The negation head-shake should already start by RAIN and continue through NOTHING |
| 32 | MOTHER SHOW BOOK DOWN-WARD | DA01: Pictures of 32.2 and 32.4 are too close. Both pictures indicate "showing to side" (Picture 32.2 depicts a boy showing book to a girl sitting next to him, 32.4 depicts the mother sitting on a chair showing the book downward to the boy sitting on the floor)<br>DA03: Same as DA01<br>DA04: Same as DA01 & DA03 | *Pictures*: 32.2 needs to indicate more clearly "shows-to-side" and 32.4 (the target) "shows-downward" |
| 34 | DOG-IN-FRONT | DA03: No clear indication that the dog is "in front of"; wrong use of classifier, should be classifier with V-Bent handshape with bent fingers<br>DA04: Classifier for dog is wrong<br>DA05: Classifier used is OK; word order should be different, should be BOX DOG CL-DOG-IN-FRONT | *Video*: Correct wrong use of classifier (V-Bent handshape) and word order to BOX CL-DOG-IN-FRONT |

| *Item No** | *Item*** | *Comments by Deaf adults* | *Researcher's conclusion/ comments and resulting revisions of stimuli* |
|---|---|---|---|
| 35 | NOT-DROP-CUP | DA01: 35.2 and 35.3 are too close, both could indicate the meaning "not drop cup" (Picture 35.2 depicts a girl holding a cup and Picture 35.3 depicts a girl that moves the cup (full of water) towards a boy standing opposite of her)<br>DA03: Same as DA01<br>DA04: Same as DA01 & DA03<br>DA05: Same as DA01, DA03, & DA04 | *Video*: Signing should more clearly indicate that the girl is holding, not dropping, the cup<br>*Pictures*: 35.2 (target) and 35.3 are too close, need to be changed |
| 38 | ROW-OF-CARS BOTTOM LEFT | DA03: Could be both 38.1 and 38.2, depending on position/perspective (Picture 38.1 depicts a row of cars that are parked closer from the viewer's perspective and Picture 38.2 depicts a row of cars that are parked further away)<br>DA04: Same as DA03<br>DA05: Same as DA03 & DA04, plus the hand orientation should be different (rotated 180 degrees) | *Video*: Needs to be signed closer to the signer's body in order to indicate the location of the row of cars<br>*Pictures*: 38.1 and 38.2; Change either picture to make distinction between location of car clearer |
| 39 | DOG-LIE-INSIDE-LEFT | DA03: Spatial reference of the dog at the side of the box is not clear; wrong classifier for dog<br>DA04: Classifier for dog is wrong | *Video*: Change classifier for dog and make clearer the relative position of the dog to the side of the box |
| 42 | PENCIL FEW | | *Pictures*: Size of all pencils on all pictures should be the same (researcher) |

| *Item No** | *Item*** | *Comments by Deaf adults* | *Researcher's conclusion/ comments and resulting revisions of stimuli* |
|---|---|---|---|
| 45 | DOG BEHIND (BOX) | DA03: Classifier for dog is wrong<br>DA04: Same as DA03<br>DA05: Word order should be first BOX, then DOG and classifier | *Video*: Use of different classifier for dog, and change word order to BOX DOG CL-DOG-BEHIND-BOX<br>Pictures: 45.3 Now looks like that dog lies beside the box; should change so dog clearly lies behind the box (researcher) |
| 46 | CHILD SWEATER STRAIGHT-ROW-DOTS | DA01: Could be a combination of 46.1 and 46.4 (Picture 46.1 depicts a girl wearing a sweater with rows of small dots, Picture 46.4 depicts the same, but the dots are larger, three in a row). The pattern on 46.1 is too small to justify the used handshape; 46.4 could use this handshape, but not with this "tracking" movement (only more than three dots would justify this "tracking" movement and a movement indicating all three dots separately)<br>DA03: Same as DA01 | *Pictures*: 46.1 enlarge dots on picture, so that they match the video |
| 47 | ONE BALL | DA03: Subject is missing; handshape is wrong, should be either a one or two handed sign with 5-Clawed handshape.<br>DA05: Same as DA03 | *Video*: Change handshape for ball to 5-Clawed handshape.<br>*Pictures*: 47.1-47.3 Make the balls on the pictures all the same size (Picture 47.1 depicts three balls, 37.2 one ball, and 37.3 depicts three rows of balls) |

| *Item No** | *Item*** | *Comments by Deaf adults* | *Researcher's conclusion/ comments and resulting revisions of stimuli* |
|---|---|---|---|
| 48 | CHILD STAND-ING-IN-FRONT-OF-CAR | DA03: The direction of the cars in the pictures is confusing, should be turned around 180 degrees so that they correspond to the signing | *Pictures*: 48.2 and 48.3: The pictures need to be rotated by 180 degrees (Picture 48.2 depicts a boy behind a car, Picture 48.3 a boy standing beside the car) |
| 49 | PENCIL SMALL | | *Pictures*: 49.3 The color of the pencil should be the same as that of the pencils in the other pictures (researcher) (All pictures depict pencils of different sizes) |

***** Number of items follows Pilot 2 (not yet) revised version (for main study)
** The pictures (answers) of the items are numbered clockwise: (1) upper left; (2) upper right; (3) lower left; and (4) lower right. For example, the picture of Item 2 in the upper right is numbered as 2.2.

## Appendix F-4
## Quantitative Results of Pilot 2 with Deaf Adults ($N = 5$)

| *Informants* | *Items* | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P1 | P2 | P3 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| #1 | 2 | 1 | 3 | 2 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 3 | 2 | 4 | 3 | 1 |
| #2 | 2 | 1 | 3 | 2 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 3 | 2 | 4 | 3 | 1 |
| #3 | 2 | 1 | 3 | 2 | 2 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 3 | 2 | 4 | 3 | 1 |
| #4 | 2 | 1 | 3 | 2 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 3 | 2 | 4 | 3 | 2 |
| #5 | 2 | 1 | 2 | 2 | 3 | 3 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 3 | 2 | 4 | 3 | 1 |
| *Targets* | | | 3 | | 3 | | | | | | | | | | | | | 1 |

| *Informants* | *Items* | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
| #1 | 3 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 2 | 4 | 2 | 2 | 4 | 1 |
| #2 | 3 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 2 | 4 | 2 | 2 | 4 | 1 |
| #3 | 3 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 2 | 4 | 3 | 2 | 2 | 1 |
| #4 | 3 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 2 | 4 | 2 | 2 | 4 | 1 |
| #5 | 3 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 4 | 3 | 4 | 2 | 4 | 2 | 2 | 2 | 1 |
| *Targets* | | | | | | | | | | | | | | | 2 | | 4 | |

| *Informants* | *Items* 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #1 | 3 | 2 | 2 | 2 | 1 | 1 | 4 | 1 | 3 | 4 | 2 | 3 | 1 | 2 | 1 | 3 | 2 |
| #2 | 2 | 2 | 2 | 4 | 1 | 1 | 4 | 1 | 3 | 4 | 2 | 4 | 1 | 2 | 1 | 3 | 2 |
| #3 | 1 | 2 | 2 | 2 | 2 | 1 | 4 | 1 | 3 | 4 | 2 | 3 | 1 | 2 | 1 | 3 | 2 |
| #4 | 3 | 2 | 2 | 3 | 1 | 1 | 4 | 1 | 3 | 4 | 2 | 1 | 4 | 2 | 2 | 3 | 2 |
| #5 | 3 | 2 | 2 | 2 | 1 | 1 | 4 | 1 | 3 | 4 | 2 | 3 | 4 | 2 | 1 | 3 | 2 |
| *Targets* | 3 | | | 2 | 1 | | | | | | | 3 | 1 | | 1 | | |

# Appendix F-5
# Cover Letter, Background Questionnaire, and Consent Form for Pilot 2 for Non-Signing Hearing Children (Translation)

Project Sign Language Tests

German Sign Language Receptive Skills Test

Dear parents,
I am very grateful that your child will participate in this pilot study, which is part of my dissertation project. I would like to take the time to thank you for your support. All data will be treated anonymously. The name of your child and your place of residence will not be mentioned in my dissertation.
The goal of my dissertation is to develop a receptive skills test for German Sign Language for Deaf children from age 4-years old. In the first pilot with hearing children ages 4–10 years old, I will investigate how non-signing hearing children are able to get the item correctly by guessing. These results are very important for the revision of the first version of the test.
I kindly ask you to answer the three questions below and to also sign the consent form. Please return the questionnaire with your child via the kindergarten.
Should you have any questions regarding my research project, please contact me at:
Email: projekt@gebaerdensprachtest.de
Cell phone: 0176 – 20 14 95 28
Fax: 069 – 79 12 500 37

1. Your child's date of birth (month/year; e.g., 07/82): ___________________

2. Did or does your child have contact with hearing-impaired people?
❑ yes
❑ no

3. Does you child have any skills in sign language?
❑ yes
❑ no

4. Comments:

Thank you very much for your collaboration!

### Consent form

Hereby, I agree that my child can participate in the pilot study of the German Sign Language Receptive Skills Test. All collected data (questionnaires, test results) can be used by Tobias Haug for his dissertation.

I also have been informed that all data will be processed anonymously.

Signature Place/date

## Appendix F-6
## Item Recoding Based on Pilot 2

### Item recoding (no changes to practice items)

| *Old item number (BSL)** | *Name old/new item* | *New item number (DGS)* |
|---|---|---|
| 1 | VIELE ÄPFEL | 1 |
| 2 | AUTO REIHE REIHE REIHE | 2 |
| 3 | EIS OHNE | 3 |
| 4 | ESSEN MÖGEN-NICHT | 4 |
| 5 | BUCH AUF (BETT) | 5 |
| 6 | EIN TEDDY | 6 |
| 7 | AUTO-FAHREN | |

| *Old item number (BSL)** | *Name old/new item* | *New item number (DGS)* |
|---|---|---|
| 8 | HUT OHNE | 7 |
| 9 | BALL TISCH-AUF | 8 |
| 10 | ZWEI-PERSONEN-TREFFEN | 9 |
| 11 | HUND IN (KISTE) | 10 |
| 12 | PERSON-ROLLTREPPE-RUNTER-FAHREN | 11 |
| 13 | KIND SCHAUT-AUF/HOCH | 12 |
| 14 | EINIGE TASSEN | 13 |
| 15 | AUTO HINTER (HAUS) | 14 |
| 16 | HAAR-LOCKIG | 15 |
| 17 | KISTE UNTER (BETT) | 16 |
| 18 | (MUTTER) BUCH-KIND-GEBEN | 17 |
| 19 | JUNGE-TRINKEN | |
| 20 | JUNGE-SCHLAGEN MÄDCHEN-WIRD-GESCHLAGEN | 18 |
| 21 | BLEISTIFT DICK | 19 |
| 22 | BREITE-STREIFEN-NACH-UNTEN HOSE | 20 |
| 23 | NICHT-SCHLAFEN | 21 |
| 24 | SCHLANGE-LEUTE | 22 |
| 25 | REGENSCHIRM-OFFEN-HALTEN GEHEN | 23 |
| 26 | BLEISTIFT | |
| 27 | (JUNGE-links) WASSER-KOPF-GIESSEN AUF (JUNGE-rechts) WASSER-KOPF-GIESSEN HAAR-NASS | 24 |
| 28 | KOPFHÖRER OHNE | 25 |
| 29 | MUTTER BRIEF GEBEN | 26 |
| 30 | KIND JACKE REGEN-NICHTS/KEIN | 27 |
| 31 | NICHT-HOCH-KOMMEN-AN-SCHRANK | 28 |
| 32 | KIND BUCH-ZEIGEN-ZUR-SEITE | 29 |

| *Old item number (BSL)** | *Name old/new item* | *New item number (DGS)* |
|---|---|---|
| 33 | HUND HALSBAND OHNE KNOCHEN-GROSS ES-SEN | 30 |
| 34 | HUND VOR (KISTE) | 31 |
| 35 | BECHER/TASSE HERUNTERFALLEN-NICHT | 32 |
| 36 | BALL OHNE | 33 |
| 37 | ESSEN-DÜNNES-SANDWICH | 34 |
| 38 | REIHE-AUTO UNTEN (LINKS) | 35 |
| 39 | HUND-LIEGEN-INNEN-LINKS (KISTE) | 36 |
| 40 | HAUS-OBEN-RECHTS | 37 |
| 41 | VIELE-AUTO | 38 |
| 42 | EINIGE-STIFTE | 39 |
| 43 | KIND-SITZEN | |
| 44 | TEDDY-UNTER-BETT | 40 |
| 45 | HUND-HINTER-KISTE | 41 |
| 46 | REIHEN-MIT-PUNKTEN-PULLI | 42 |
| 47 | EIN-BALL | 43 |
| 48 | KIND-VOR-AUTO-STEHEN | 44 |
| 49 | STIFT-KLEIN | 45 |
| 50 | JUNGE-GETRÄNK-OHNE | 46 |

* I1-40 BSL test, I41-50 newly developed DGS items

# Appendix G-1
# Parent Questionnaire for Main Study (Translation)

Dear parent,
Thank you for participating in this survey. Your answers will be treated completely anonymously. Please send the completed questionnaire back with your child or fax it to the following number: 069 – 79 12 500 37. If you have any further questions regarding the project, please contact me, Tobias Haug, at 0176 – 20 14 95 28 or by email (projekt@gebaerdensprachtest.de).

1. Child's ID: ______________________________

2. Child's date of birth (month/year; e.g., 7/82): ________________

3. How old was the child when the hearing loss was diagnosed?
   ________ years, ________months

4. How old was your child when she/he started to learn sign language?
   ________ years, ________months

5. What language does your family most often use at home?
   - ❑ German
   - ❑ German Sign Language (DGS)
   - ❑ Signed German
   - ❑ Home made signs/gestures
   - ❑ Other (specify):

6. What is your first language (if not German)?

   ______________________________

   6.1 What is your partner's/spouse's first language (if not German)?

   ______________________________

7. Please tell us about all the people (including yourself) who live in your home by listing them below.

| *Relationship to child (e.g., mother or father)* | *Deaf or hearing?* | *Languages used to communicate with the child* |
|---|---|---|
| 1. | | |
| 2. | | |
| 3. | | |
| 4. | | |
| 5. | | |
| 6. | | |

8. Does the Deaf child have contact with any person not listed above (outside of school) who knows sign language?
   ___ yes (please answer Question 8.1)
   ___ no (please skip to Question 9.)
   8.1 If yes, please tell us about these people by listing them below:

| *Relationship to child* | *Deaf or hear-ing?* | *How long has this person known sign language?* | *How often are these people in contact with your child?* |
|---|---|---|---|
| 1. | | | |
| 2. | | | |
| 3. | | | |
| 4. | | | |
| 5. | | | |
| 6. | | | |

9. What language does the Deaf child use most frequently at home? (please check only one)
   ❑ German
   ❑ German Sign Language (DGS)
   ❑ Signed German
   ❑ Home made signs/gestures
   ❑ Other (please specify): ______________________________

   Thank you for your participation in this questionnaire!

## Appendix G-2
## Student Questionnaire (Through Teachers) (Translation)

Dear colleagues,
Thank you for your participation in this study. The provided information is important for a meaningful analysis of students' test results. All answers will be treated completely anonymously.

1. ID of participating child: ______________________________

2. Date of birth (month/year, e.g. 07/82): ______________________

3. Date of child's enrollment: ______________________________
   3.1 Do you know at what age (e.g., 3 years) the child started to use sign language?
   If yes, please provide the approximate age: __________________
   3.2 Please indicate the child's degree of hearing loss:
   - ❑ mild (25 to 40 dB)
   - ❑ moderate (40 to 70 dB)
   - ❑ severe (70 to 90 dB)
   - ❑ profound (> 100 dB)

   3.3 When was the child's hearing loss diagnosed?
   ________years__________months

4. Hearing status of parents:
   Mother
   ❑ Deaf ❑ hard of hearing ❑ hearing
   Father
   ❑ Deaf ❑ hard of hearing ❑ hearing

5. What form of communication is used in the child's home, to your knowledge? (please indicate all that apply)
   - ❑ German
   - ❑ German Sign Language
   - ❑ Signed German
   - ❑ Home signs
   - ❑ Other: ______________________________

6. To the best of your knowledge, does the child have contact with people outside of school who sign (e.g., uncle, friend)?
   - ❑ yes
   - ❑ no

7. Please rate the child's DGS skills (i.e., comprehension and production) using the following scales (1=highest, 6=lowest).

*DGS Comprehension:*

1 2 3 4 5 6

Comments: ______________________________________________

*DGS Production:*

1 2 3 4 5 6

Comments: ______________________________________________

# Appendix G-3
# Teacher Questionnaire (Translation)

Dear colleague,
Thank you for your participation in this study. Your answers will be treated completely anonymously. You can drop off the completed questionnaire at the principal's office or send it directly by fax.

## I. Demographics

1. I work in my school as a:
   - ❑ Teacher
   - ❑ Speech Language Pathologist
   - ❑ School Psychologist
   - ❑ Other (please specify) ________________________________

2. What is your hearing status?
   - ❑ Hearing
   - ❑ Hard of Hearing
   - ❑ Pre-lingual Deaf
   - ❑ Post-lingual Deaf
   - ❑ Other (please specify) ____________________

3. What classes/year do you teach that have Deaf students?

4. Does your school use sign language as means of instruction in class?
   ___ yes (please answer Question 4.1)
   ___ no (please skip to Question 5.)
   If you answered yes, please indicate the means of communication and the frequency of use.

| | *Frequently* | *Often* | *Sometimes* | *Rarely* | *Never* |
|---|---|---|---|---|---|
| *Spoken language* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *German Sign Language* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *Signed German* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *Finger spelling* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *Lip reading* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *Home signs* | ❑ | ❑ | ❑ | ❑ | ❑ |
| *Other* | ❑ | ❑ | ❑ | ❑ | ❑ |

## II. Communication

5. Which of the following means of communication are used…

| | *Spoken language* | *DGS* | *Signed German* | *Finger spelling* | *Lip-reading* | *Home sign* | *Other* |
|---|---|---|---|---|---|---|---|
| *5.1 … by you during class?* | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| *5.2 … by you outside of class?* | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| *5.3 … between the students during class?* | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| *5.4 … between the students during recess?* | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |

For other, please specify:

6. Please rate your own DGS skills (i.e., comprehension and production) using the following scales.

| *Sign language perception* | *Sign language production* |
|---|---|
| Score 1: I can comprehend several signs and simple sentences when they are signed slowly and with repetitions. | Score 1: I can produce a few signs (slowly) and reply to basic questions. |
| Score 2: I can comprehend basic/simple signed sentences, but I frequently have to ask to be able to follow a conversation in sign language. | Score 2: I can produce basic sentences (slowly), but I often have to think about how to express my thoughts/ideas in sign language. |
| Score 3: I feel quite confident about following a conversation in sign language, but occasionally I have to ask in order to understand everything. | Score 3: I feel quite confident about participating in a conversation in sign language, but occasionally I have to think about how to express my thoughts in signs. |
| Score 4: I can understand/follow almost all conversations in sign language. | Score 4: I can participate confidently in almost all conversations in sign language. |
| Score 5: I am able to comprehend conversations in sign language on any topic. | Score 5: I am able to participate actively in conversations in sign language on any topic. |
| ❑ I don't use sign language | |

7. How many students are in your class?
Number of students: __________
7.1. How many of these students are...
____ Deaf _____ hard of hearing _____ CI users
7.2. How many of the Deaf students have an additional disability?
Number of students: ___________

## Appendix G-4
## Observation Sheet Used During Testing (Test Administrator) (Translation)

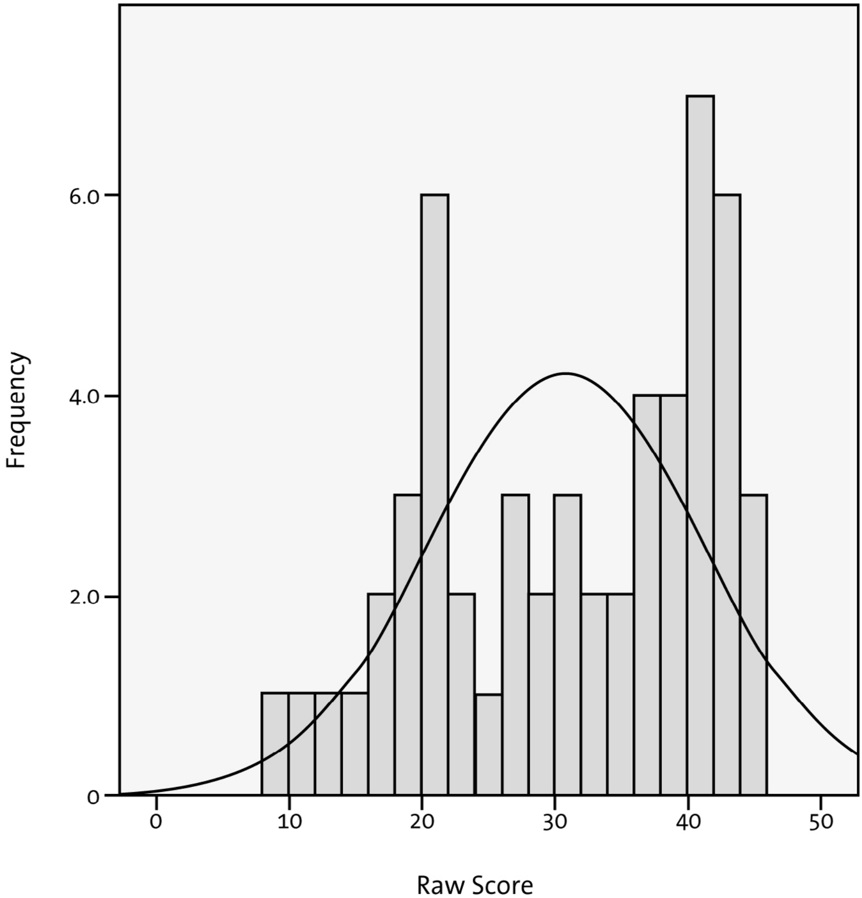
German Sign Language Receptive Skills Test (DGS-VT)
Observation sheet

Date & time: ________________________________________
ID of child: ________________________________________
Gender: ________________________________________
Test administrator: ________________________________________
General procedure

1. Test administrator introduces himself and explains what is expected of the child (watch films and videos on the computer)
2. Child introduces himself/herself, including name sign
3. Information that whole session will be video-taped

1. Does the child have experience with a computer?
   - ❑ yes
   - ❑ no
2. Can the child use the mouse himself/herself?
   - ❑ yes
   - ❑ no → Continue with Question 3
3. In order to select the right answer, the child points …
   - ❑ to the print-outs of the pictures in the booklet
   - ❑ to/on the screen
   - ❑ a combination of both
4. Observations

# Appendix H-1
# Histogram Raw Score with Normal Curve Overlaid



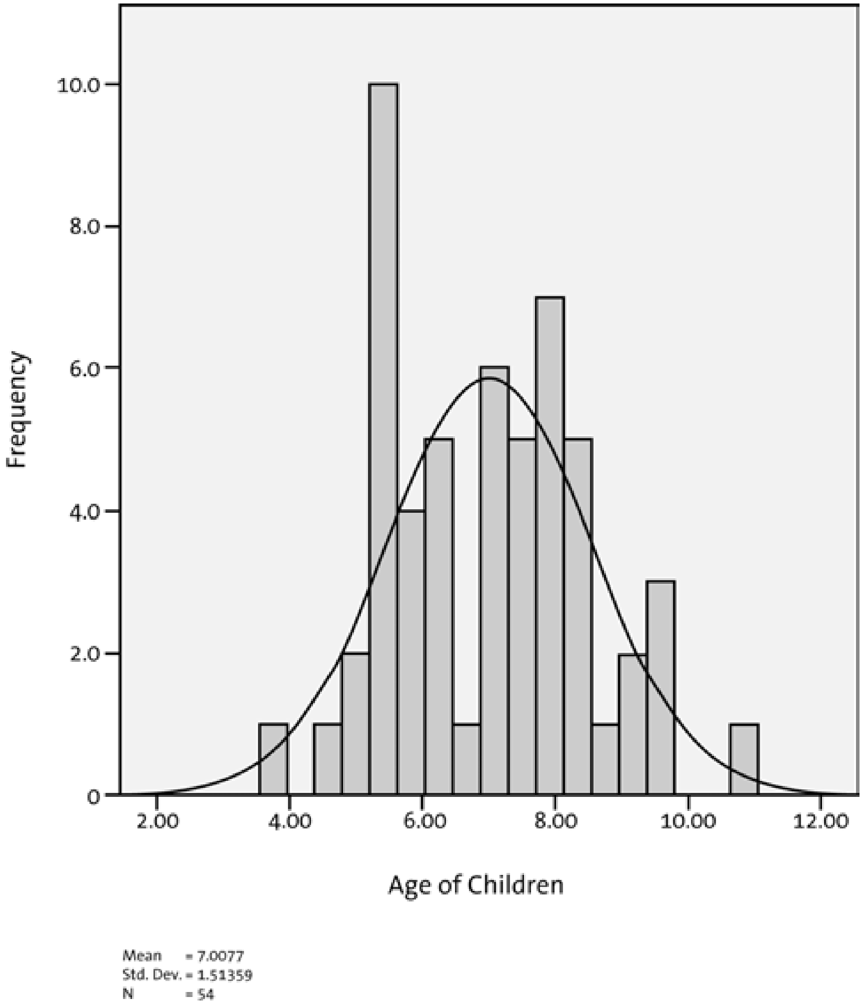

Mean = 30.72
Std. Dev. = 10.151
N = 54

# Appendix H-2
# Descriptive Statistics of the Variable Raw Score

Descriptive statistics of the variable Raw Score (*N* = 54)

| *Statistics* | | |
|---|---|---|
| Mean | | 30.72 |
| Standard Error of Mean | | 1.38 |
| Standard Deviation | | 10.15 |
| Variance | | 103.03 |
| Minimum | | 9 |
| Maximum | | 44 |
| Percentiles | 10 | 16.5 |
| | 20 | 20 |
| | 30 | 22.5 |
| | 40 | 28 |
| | 50 | 32.5 |
| | 60 | 37 |
| | 70 | 39.5 |
| | 80 | 40 |
| | 90 | 42 |

## Appendix H-3
## Histogram Age with Normal Curve Overlaid



# Appendix H-4
# Normal Q-Q Plot of Age

# Appendix I-1
# Results of Item Analysis of Deaf Children of Deaf Parents

Item analysis of Deaf children of Deaf parents ($N$ = 34)

| *Item No. DGS Pilot 2* | *Item No. DGS main study* | *Facility value $p_i$* | *Item discrimination $r_{it}$* | *Discarded=D Retained=R* |
|---|---|---|---|---|
| I41 | I38 | .971 | .396* | D |
| I9 | I8 | .882 | .610** | R |
| I8 | I7 | .882 | .471** | R |
| I4 | I4 | .882 | .443** | R |
| I49 | I45 | .882 | .323 | R |
| I2 | I2 | .882 | .105 | D |
| I5 | I5 | .853 | .583** | R |
| I38 | I35 | .853 | .188 | D |
| I1 | I1 | .824 | .704** | R |
| I16 | I15 | .824 | .664** | R |
| P2 | P2 | .824 | .656** | R |
| P3 | P3 | .824 | .474** | R |
| I10 | I9 | .794 | .728** | R |
| I25 | I23 | .794 | .659** | R |
| I28 | I25 | .794 | .599** | R |
| I31 | I28 | .794 | .584** | R |
| I18 | I17 | .794 | .569** | R |
| I42 | I39 | .765 | .690** | R |
| I22 | I20 | .765 | .404* | R |
| P1 | P1 | .765 | .719** | R |
| I47 | I43 | .735 | .653** | R |
| I6 | I6 | .735 | .591** | R |
| I33 | I30 | .735 | .577** | R |
| I50 | I46 | .735 | .487** | R |
| I20 | I18 | .735 | .426* | R |

| *Item No. DGS Pilot 2* | *Item No. DGS main study* | *Facility value $p_i$* | *Item discrimination $r_{it}$* | *Discarded=D Retained=R* |
|---|---|---|---|---|
| I3 | I3 | .706 | .758** | R |
| I24 | I22 | .706 | .677** | R |
| I14 | I13 | .706 | .549** | R |
| I12 | I11 | .706 | .542** | R |
| I27 | I24 | .676 | .736** | R |
| I44 | I40 | .676 | .736** | R |
| I11 | I10 | .676 | .730** | R |
| I21 | I19 | .676 | .683** | R |
| I36 | I33 | .676 | .644** | R |
| I13 | I12 | .676 | .637** | R |
| I17 | I16 | .647 | .791** | R |
| I30 | I27 | .647 | .361* | R |
| I32 | I29 | .647 | .336 | R |
| I35 | I32 | .618 | .491** | R |
| I23 | I21 | .618 | .299 | R |
| I46 | I42 | .588 | .167 | D |
| I39 | I36 | .559 | -.310 | D |
| I29 | I26 | .529 | .654** | R |
| I37 | I34 | .324 | .271 | R |
| I15 | I14 | .235 | -.266 | D |
| I34 | I31 | .206 | .038 | D |
| I48 | I44 | .118 | -.078 | D |
| I45 | I41 | .059 | -.220 | D |
| I40 | I37 | .029 | .103 | D |

DGS Pilot 2 Numbering: P1-P3 and I1-40 are based on BLS test, I41-50 are newly developed items
DGS Main Study Numbering: P1-3 and Items 1-37 are based on the BSL test (but with different numbering), Items 38-46 are newly designed items
*correlation is significant at the 0.05 level (2-tailed)
** correlation is significant at the 0.01 level (2-tailed)

# Appendix I-2
# Distractor Analysis of Deaf Children of Deaf Parents

| *Item* | *Answer 1 / Frequency* | *Facility value 1* | *Dis. index d1* | *Answer 2 / Frequency* | *Facility value 2* | *Dis. index d2* | *Answer 3 / Frequency* | *Facility value 3* | *Dis. index d3* | *Answer 4 / Frequency* | *Facility value 4* | *Dis. index d4* | *Score* | *Missing value* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 4 | 0.1176 | -0.477** | 26 | 0.7647 | | 4 | 0.1176 | -0.495** | N/A | N/A | N/A | 34 | |
| P2 | 28 | 0.8235 | | 4 | 0.1176 | -0.522** | 2 | 0.0588 | -0.382* | N/A | N/A | N/A | 34 | |
| P3 | 1 | 0.0294 | -0.412* | 5 | 0.1471 | -0.345* | 28 | 0.8235 | | 0 | 0 | a | 34 | |
| I1 | 0 | 0.0000 | a | 28 | 0.8235 | | 3 | 0.0882 | -0.583** | 3 | 0.0882 | -0.389* | 34 | |
| I2# | 3 | 0.0882 | -0.032 | 1 | 0.0294 | -0.206 | 30 | 0.8824 | | 0 | 0 | a | 34 | |
| I3 | 5 | 0.1471 | -0.320 | 5 | 0.1471 | -0.680** | 24 | 0.7059 | | N/A | N/A | N/A | 34 | |
| I4 | 30 | 0.8824 | | 1 | 0.0294 | -0.103 | 3 | 0.0882 | -0.470** | 0 | 0 | a | 34 | |
| I5 | 1 | 0.0294 | -0.206 | 29 | 0.8529 | | 4 | 0.1176 | -0.558** | N/A | N/A | N/A | 34 | |
| I6 | 25 | 0.7353 | | 6 | 0.1765 | -0.632** | 3 | 0.0882 | -0.113 | N/A | N/A | N/A | 34 | |
| I7 | 2 | 0.0588 | -0.222 | 2 | 0.0588 | -0.456** | 30 | 0.8824 | | 0 | 0 | a | 34 | |
| I8 | 30 | 0.8824 | | 2 | 0.0588 | -0.542** | 2 | 0.0588 | -0.321 | 0 | 0 | a | 34 | |
| I9 | 4 | 0.1176 | -0.603** | 2 | 0.0588 | -0.419* | 1 | 0.0294 | -0.052 | 27 | 0.7941 | | 34 | |
| I10 | 7 | 0.2059 | -0.531** | 4 | 0.1176 | -0.432* | 23 | 0.6765 | | N/A | N/A | N/A | 34 | |
| I11 | 2 | 0.0588 | -0.062 | 24 | 0.7059 | | 6 | 0.1765 | -0.624** | 2 | 0.0588 | -0.038 | 34 | |
| I12 | 4 | 0.1176 | -0.612** | | 0.0882 | -0.246 | 4 | 0.1176 | -0.136 | 23 | 0.6765 | | 34 | |
| I13 | 4 | 0.1176 | -0.396* | 6 | 0.1765 | -0.358* | 24 | 0.7059 | | N/A | N/A | N/A | 34 | |
| I14# | 8 | 0.2353 | | 20 | 0.5882 | 0.461** | 5 | 0.1471 | -0.287 | 1 | 0.0294 | -0.172 | 34 | |
| I15 | 3 | 0.0882 | -0.440** | 2 | 0.0588 | -0.284 | 28 | 0.8235 | | 1 | 0.0294 | -0.412* | 34 | |
| I16 | 5 | 0.1471 | -0.484** | 22 | 0.6471 | | 7 | 0.2059 | -0.531** | 0 | 0 | a | 34 | |
| I17 | 1 | 0.0294 | -0.292 | 6 | 0.1765 | -0.503** | 0 | 0.0000 | a | 27 | 0.7941 | | 34 | |
| I18 | 25 | 0.7353 | | 0 | 0 | a | 7 | 0.2059 | -0.381* | 2 | 0.0588 | -0.210 | 34 | |
| I19 | 1 | 0.0294 | -0.292 | 23 | 0.6765 | | 3 | 0.0882 | -0.593** | 7 | 0.2059 | -0.281 | 34 | |
| I20 | 26 | 0.7647 | | 5 | 0.1471 | -0.524** | 1 | 0.0294 | -0.206 | 2 | 0.0588 | 0.147 | 34 | |
| I21 | 21 | 0.6176 | | 8 | 0.2353 | -0.240 | 5 | 0.1471 | -0.181 | N/A | N/A | N/A | 34 | |
| I22 | 0 | 0.0000 | a | 24 | 0.7059 | | 3 | 0.0882 | -0.450** | 7 | 0.2059 | -0.474** | 34 | |

| Item | Answer 1 / Frequency | Facility value 1 | Dis. index d1 | Answer 2 / Frequency | Facility value 2 | Dis. index d2 | Answer 3 / Frequency | Facility value 3 | Dis. index d3 | Answer 4 / Frequency | Facility value 4 | Dis. index d4 | Score | Missing value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I23 | 4 | .1176 | -.352* | 0 | .0000 | a | 3 | .0882 | -.573** | 27 | .7941 | | 34 | |
| I24 | 10 | .2941 | -.625** | 1 | .0294 | -.412* | 0 | .0000 | a | 23 | .6765 | | 34 | |
| I25 | 3 | .0882 | -.521** | 27 | .7941 | | 4 | .1176 | -.325 | N/A | N/A | N/A | 34 | |
| I26 | 12 | .3529 | -.323 | 2 | .0588 | -.456** | 2 | .0588 | -.333 | 18 | .5294 | | 34 | |
| I27 | 5 | .1471 | -.107 | 22 | .6471 | | 2 | .0588 | -.308 | 5 | .1471 | -.230 | 34 | |
| I28 | 2 | .0588 | -.456** | 27 | .7941 | | 5 | .1471 | -.394* | N/A | N/A | N/A | 34 | |
| I29 | 3 | .0882 | -.379* | 4 | .1176 | -.064 | 5 | .1471 | .148 | 22 | .6471 | | 34 | |
| I30 | 25 | .7353 | | 1 | .0294 | -.206 | 4 | .1176 | -.576** | 4 | .1176 | -.145 | 34 | |
| I31# | 16 | .4706 | .357* | 8 | .2353 | -.275 | 7 | .2059 | | 3 | .0882 | -.328 | 34 | |
| I32 | 7 | .2059 | -.180 | 21 | .6176 | | 3 | .0882 | -.215 | 3 | .0882 | -.430* | 34 | |
| I33 | 1 | .0294 | -.206 | 23 | .6765 | | 2 | .0588 | -.468** | 8 | .2353 | -.397* | 34 | |
| I34 | 8 | .2353 | -.206 | 11 | .3235 | | 7 | .2059 | .220 | 8 | .2353 | -.350* | 34 | |
| I35# | 29 | .8529 | | 4 | .1176 | -.190 | 1 | .0294 | -.103 | 0 | .0000 | a | 34 | |
| I36# | 19 | .5588 | | 2 | .0588 | -.136 | 2 | .0588 | -.062 | 11 | .3235 | .382* | 34 | |
| I37# | 7 | .2059 | -.037 | 16 | .4706 | .432* | 10 | .2941 | -.485** | 1 | .0294 | | 34 | |
| I38# | 33 | .9706 | | 0 | .0000 | a | 1 | .0294 | -.412* | 0 | .0000 | a | 34 | |
| I39 | 1 | .0294 | -.103 | 3 | .0882 | -.215 | 26 | .7647 | | 4 | .1176 | -.693** | 34 | |
| I40 | 0 | .0000 | a | 23 | .6765 | | 10 | .2941 | -.745** | 0 | .0000 | a | 33 | 1 |
| I41# | 25 | .7353 | .350* | 3 | .0882 | -.174 | 2 | .0588 | | 3 | .0882 | -.154 | 33 | 1 |
| I42# | 20 | .5882 | | 5 | .1471 | -.402* | 2 | .0588 | -.210 | 6 | .1765 | .264 | 33 | 1 |
| I43 | 3 | .0882 | -.154 | 25 | .7353 | | 5 | .1471 | -.680** | N/A | N/A | N/A | 33 | 1 |
| I44# | 4 | .1176 | | 24 | .7059 | .555** | 2 | .0588 | -.198 | 3 | .0882 | -.624** | 33 | 1 |
| I45 | 1 | .0294 | -.206 | 2 | .0588 | -.271 | 30 | .8824 | | N/A | N/A | N/A | 33 | 1 |
| I46 | 4 | .1176 | -.423* | 25 | .7353 | | 4 | .1176 | -.244 | N/A | N/A | N/A | 33 | 1 |

Legends
\# = items that need to be rivised/removed based on items analysis
0 = distractor was not chosen
N/A = item has only three MC answers
** = correlation is significant at 0.01 level (2-tailed)
* = correlation is significant at 0.05 level (2-tailed)
a = can't be computed because one of the variables is constant

# Appendix I-3
# Homogeneity Indices *H* of All Test Items

| *Items* | *H index/item* |
|---|---|
| P1 | .444 |
| P2 | .412 |
| P3 | .287 |
| I1 | .420 |
| I3 | .464 |
| I4 | .244 |
| I5 | .348 |
| I6 | .353 |
| I7 | .302 |
| I8 | .343 |
| I9 | .446 |
| I10 | .449 |
| I11 | .346 |
| I12 | .381 |
| I13 | .363 |
| I15 | .411 |
| I16 | .479 |
| I17 | .348 |
| I18 | .275 |
| I19 | .415 |
| I20 | .243 |
| I21 | .197 |
| I22 | .423 |
| I23 | .401 |

| *Items* | *H index/item* |
|---|---|
| I24 | .455 |
| I25 | .364 |
| I26 | .405 |
| I27 | .201 |
| I28 | .341 |
| I29 | .215 |
| I30 | .349 |
| I32 | .303 |
| I33 | .384 |
| I34 | .162 |
| I39 | .423 |
| I40 | .455 |
| I43 | .405 |
| I45 | .201 |
| I46 | .302 |
| *H total* | *.353* |

# Appendix I-4
# Linear Regression Model of Deaf Children of Hearing Parents (*N* = 20)

## Model Summary

| *R* | *R Square* | *Adjusted R Square* | *Std. Error of the Estimate* |
|---|---|---|---|
| .563 | .317 | .279 | 7.452 |

The independent variable is Age in Years

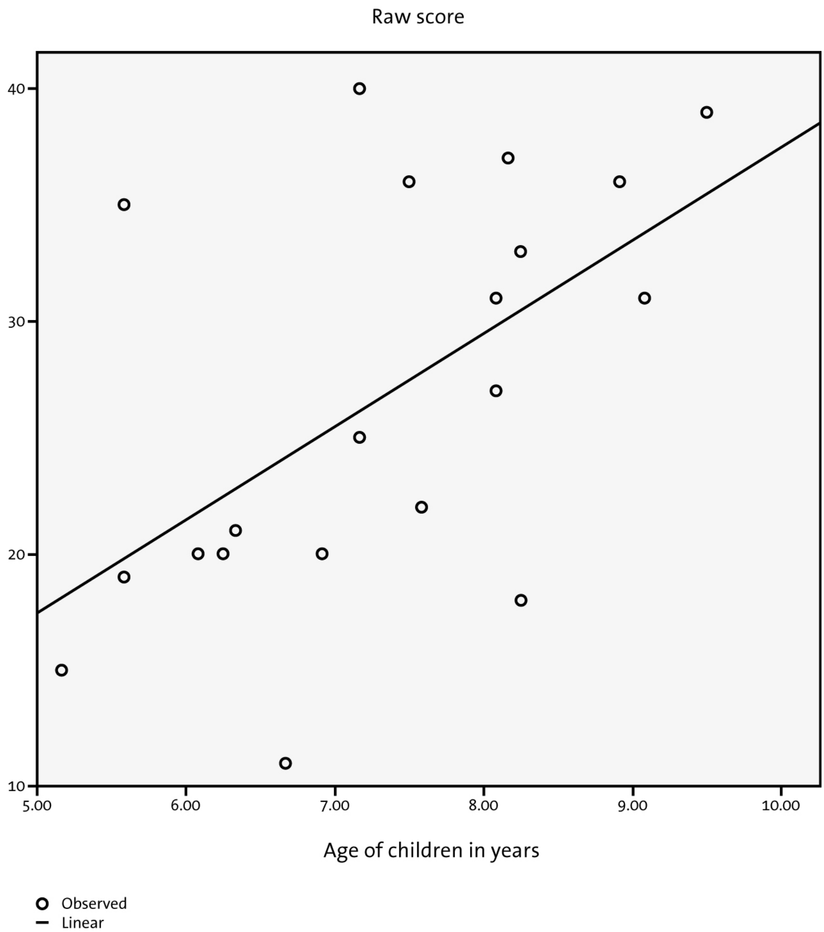
## ANOVA

| | *Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|
| Regression | 463.557 | 1 | 463.557 | 8.347 | .010 |
| Residual | 999.643 | 18 | 55.536 | | |
| Total | 1463.200 | 19 | | | |

The independent variable is Age in Years

## Coefficients

| | *Unstandardized Coefficients* | | *Standardized Coefficients* | | |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. |
| *Age in Years* | 3.993 | 1.382 | .563 | 2.889 | .010 |
| *(Constant)* | -2.414 | 10.248 | | -.236 | .816 |


Raw score
40
30
20
10
5.00
6.00
7.00
8.00
9.00
10.00
Age of children in years
Observed
Linear

# Appendix I-5
# Linear Regression Model of Deaf Children of Deaf Parents (*N* = 34)

## Model Summary

| *R* | *R Square* | *Adjusted R Square* | *Std. Error of the Estimate* |
|---|---|---|---|
| .636 | .405 | .386 | 8.080 |

The independent variable is Age in Years

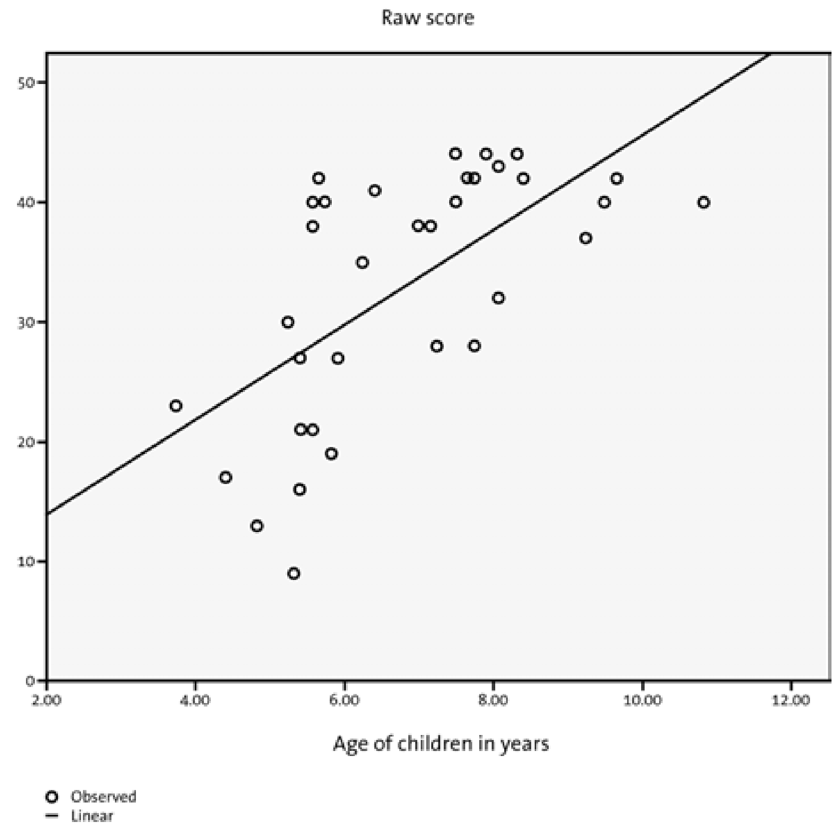
## ANOVA

| | *Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|
| *Regression* | 1419.770 | 1 | 1419.770 | 21.746 | .000 |
| *Residual* | 2089.201 | 32 | 65.288 | | |
| *Total* | 3508.971 | 33 | | | |

The independent variable is Age in Years

## Coefficients

| | *Unstandardized Coefficients* | | *Standardized Coefficients* | | |
|---|---|---|---|---|---|
| | *B* | *Std. Error* | *Beta* | *t* | *Sig.* |
| *Age in Years* | 3.987 | .855 | .636 | 4.663 | .000 |
| *(Constant)* | 5.814 | 5.998 | | .969 | .340 |


Raw score
50
40
30
20
10
0
2.00
4.00
6.00
8.00
10.00
12.00
Age of children in years
Observed
Linear

# Appendix I-6
# Regression Model with Logistic Curve Fit of Deaf Children of Hearing Parents (*N* = 20)

## Model Summary

| *R* | *R Square* | *Adjusted R Square* | *Std. Error of the Estimate* |
|---|---|---|---|
| .553 | .306 | .267 | .673 |

The independent variable is Age in Years

## ANOVA

| | *Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|
| Regression | 3.590 | 1 | 3.590 | 7.937 | .011 |
| Residual | 8.143 | 18 | .452 | | |
| Total | 11.733 | 19 | | | |

The independent variable is Age in Years

## Coefficients

| | *Unstandardized Coefficients* | | *Standardized Coefficients* | | |
|---|---|---|---|---|---|
| | *B* | *Std. Error* | *Beta* | *t* | *Sig.* |
| *Age in Years* | .704 | .088 | .575 | 8.017 | .000 |
| *(Constant)* | .215 | .199 | | 1.081 | .294 |

The dependent variable is Raw Score

# Appendix I-7
# Regression Model with Logistic Curve Fit of Deaf Children of Deaf Parents ($N$ = 34)

## Model Summary

| *R* | *R Square* | *Adjusted R Square* | *Std. Error of the Estimate* |
|---|---|---|---|
| .631 | .398 | .380 | .808 |

The independent variable is Age in Years

## ANOVA

| | *Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|
| *Regression* | 13.828 | 1 | 13.828 | 21.186 | .000 |
| *Residual* | 20.887 | 32 | .653 | | |
| *Total* | 34.715 | 33 | | | |

The independent variable is Age in Years

## Coefficients

| | *Unstandardized Coefficients* | | *Standardized Coefficients* | | |
|---|---|---|---|---|---|
| | *B* | *Std. Error* | *Beta* | *t* | *Sig.* |
| *Age in Years* | .968 | .007 | .532 | 140.369 | .000 |
| *(Constant)* | .126 | .076 | | 1.667 | .105 |

The dependent variable is Raw Score

## Appendix J-1
## All Test Items, including Revisions Which Would Be Necessary for a Standardization Study

Revisions to items for standardization study

| *Order of items based on level of difficulty (easier items first)* | *Results of item analysis* | *Revisions / changes* | |
|---|---|---|---|
| *Item number DGS Main Study* | *Discarded=D Retained=R* | *Revisions of distractors (pictures)* | *New sampling / comments* |
| I38 | D | | Linguistic concept is already represented in a different item, no need for revision. |
| I8 | R | Passed item analyses, one distractor (8.4) not chosen at all, suggestion to revise distractor. | |
| I7 | R | Passed item analyses, one distractor (7.4) not chosen at all, suggestion to revise distractor. | |
| I4 | R | Passed item analyses, one distractor (4.4) not chosen at all, suggestion to revise distractor. | |
| I45 | R | | |

| *Order of items based on level of difficulty (easier items first)* | *Results of item analysis* | *Revisions / changes* | |
|---|---|---|---|
| I2 | D | | Failed item analysis, one distractor (2.4) not chosen at all; because the linguistic concept is not represented in another item. it is suggested for new sampling. |
| I5 | R | | |
| I35 | D | | Failed item analysis, one distractor (35.4) not chosen at all; because the linguistic concept is not represented in another item, it is suggested for new sampling. |
| I1 | R | Passed item analyses, one distractor (1.1) not chosen at all, suggestion to revise distractor. | |
| I15 | R | | |
| P2 | R | | |
| P3 | R | Passed item analyses, one distractor (P3.4) not chosen at all, suggestion to revise distractor. | |
| I9 | R | | |
| I23 | R | Passed item analyses, one distractor (23.2) not chosen at all, suggestion to revise distractor. | |
| I25 | R | | |
| I28 | R | | |

| *Order of items based on level of difficulty (easier items first)* | *Results of item analysis* | *Revisions / changes* |
|---|---|---|
| I17 | R | Passed item analyses, one distractor (17.3) not chosen at all, suggestion to revise distractor. |
| I39 | R | |
| I20 | R | Passed item analysis, but one distractor showed a positive correlation (20.4). i.e. the distractor was picked more often by subjects with an overall high score than by subjects with over all lower scores (whereas a negative correlation would be expected). Therefore revise distractor and carry out new sampling. |
| P1 | R | |
| I43 | R | |
| I6 | R | |
| I30 | R | |
| I46 | R | |
| I18 | R | |
| I3 | R | |
| I22 | R | Passed item analyses, one distractor (22.1) not chosen at all, suggestion to re- |

| *Order of items based on level of difficulty (easier items first)* | *Results of item analysis* | *Revisions / changes* |
|---|---|---|
| | | vise distractor. |
| I13 | R | |
| I11 | R | |
| I24 | R | Passed item analyses, one distractor (24.3) not chosen at all, suggestion to revise distractor. |
| I40 | R | Passed item analyses, two distractors (40.1, 40.4) not chosen at all, suggestion to revise distractor. |
| I10 | R | |
| I19 | R | |
| I33 | R | |
| I12 | R | |
| I16 | R | Passed item analyses, one distractor (16.4) not chosen at all, suggestion to revise distractor. |
| I27 | R | |
| I29 | R | |
| I32 | R | |

| *Order of items based on level of difficulty (easier items first)* | *Results of item analysis* | *Revisions / changes* |
|---|---|---|
| I31 | D | Failed item analysis, one distractor (31.1) was picked over target. Because of linguistic concept, that is not represen-ted in another item, this items should undergo new sampling/piloting. |
| I44 | D | Failed item analysis, one distractor (44.2) was picked over target. Because the linguistic concept is not represented in another item, this item should undergo new sampling/piloting. |
| I41 | D | Failed item analysis, one distractor (41.1) was picked over target. Because the linguistic concept is not represented in another item, this item should undergo new sampling/piloting. |
| I37 | D | Failed item analysis, one distractor (37.2) was picked over target. Because the linguistic concept is not represented in another item, this item should undergo new sampling/piloting. |

# Appendix J-2
# Steps and Procedures for Adaptation of Sign Language Tests

## Steps and procedures for adaptation of sign language tests

| *Steps* | *Description* | *Procedure / Method* |
|---|---|---|
| (1) Test items | (a) Identify test items of source language test and check their "equivalence" in target language | (a) – (e) Review of research literature & consultations with Deaf and hearing experts (linguists, native signers) |
| | (b) Identify vocabulary of age target group | |
| | (c) Identify regional variations of different vocabulary items | (c) Small piloting (if needed) |
| | (d) Identify target language-specific structures | |
| | (e) Identify developmental pattern of structures represented in test (e.g., review of acquisition studies of other sign languages) | |
| | (f) Check suitability of distractors from the content side (not pictures, but what they represent; i.e., if they also represent a phonological, lexical, or morpho-syntactic distractor in the target language) | |
| (2) Test materials | (a) Test materials: check for cultural appropriateness of test materials and concept representation, also for distractors | (a) Consultations with Deaf and hearing experts |
| | (b) Scoring sheets: check for appropriateness of scoring sheets for the target language (e.g., certain categories do not apply in target language | |
| (3) Construct definition | (a) Defining the construct (e.g., language development) | Results of literature review of steps (1) and (2) |
| | (b) Ranking of item complexity | Results of consultations of steps (1) and (2) |

| *Steps* | *Description* | *Procedure / Method* |
|---|---|---|
| (4) Operationalization of construct (adaptation of test items) | (a) Adaptation of source language items into the target sign language<br>(b) Translation/adaptation of test instructions<br>(c) Develop additional items for the target sign language<br>(d) Decision of items order (based on ranking of item order and item order of source language test) | Results of previous steps (1) – (3) |
| (5) Technical realization | (a) Filming of adapted test items<br>(b) Decision and realization of test format (i.e., computer- or web-based test format or on DVD/video) | |
| (6) Pre-pilot test | (a) Review of pre-pilot test version<br>(b) Revisions – depending on (1) | (a) Panel of experts providing input |
| (7) Pilot with Deaf adults | (a) Conduct pilot study of the test with Deaf adults<br>(b) Revisions to test structure, items, and materials | (a) Testing<br>(a) Open-worded questionnaires/interviews for input |
| (8) Pilot test version | (a) Review of pilot by (Deaf) sign linguist(s) | (a) Prepared criteria for review |
| (9) Pilot with Deaf children | (a) Conduct pilot study with Deaf children<br>(b) Obtain social-demographic information about the Deaf children and the teachers | (a) Testing<br>(b) Questionnaires/ interviews |
| (10) Analyses of pilot study | (a) Analyze the results of the pilot study with Deaf children<br>(b) Check the effectiveness of the items<br>(c) Check the effectiveness of the distractors<br>(d) Check how well other variables explain differences in performance<br>(e) Reliability | (a) – (f) Statistical analysis (item and distractor analysis, different variables in relation to test performance) |

| *Steps* | *Description* | *Procedure / Method* |
|---|---|---|
| | (f) Validity (e.g., external variable)<br>(g) Suggestions for revisions | |
| (11) First test version | (a) Revision of pilot test version based on the results of pilot | |
| (12) Construct validation | (a) Deaf adults rank the ascribed age of acquisition of linguistic structures represented in items of first test version<br>(b) This ranking should be compared with the original ranking of item order (see construct definition) and results of pilot (item analysis) as a source of validity<br>(c) The results of (a) and (b) should be comparable | (a) Map of Ranking of Items Complexity |
| (13) Pilot with hearing children (optional) | (a) Conduct pilot study with same aged hearing non-signing peers to investigate the effect of iconicity<br>(b) Consequences of results? | (a) Testing<br>(a) Questionnaires |
| (14) Standardization study | (a) Define criteria for norming sample (e.g., linguistic experience, age groups)<br>(b) Conduct a standardization study | (b) Testing<br>(b) Questionnaires / interviews |
| (15) Analysis of standardization study | (a) Analyzing the results of standardization study<br>(b) Establishment of age-related norm<br>(c) Final revisions if needed | (a) & (b) Statistical analysis (same as above, plus standard scores) |
| (16) Publication | (a) Prepare for publication for practitioners in schools (manual, ease of test use etc.)<br>(b) Decide on the format how the test should be delivered (e.g., DVD/video, computer- or web-based) | |
| (17) Stay in touch | (a) Stay in contact with schools, further obtain information of scoring (i.e., collect scoring sheets) | |

## About the Author

Tobias Haug was born in Germany in 1971. He studied Sign Language Linguistics at Hamburg University and Deaf Education at Boston University, USA, where he received his Master's degree in 1998. In 2009, he completed his PhD in the area of sign language evaluation at Hamburg University. Since 2004 he is the co-director of the Sign Language Interpreter program at the University of Applied Sciences for Special Needs Education Zurich (HfH), Switzerland, where he also teaches.

His main research focus includes interpreting related issues and sign language evaluation/assessment.

Despite the current need for reliable and valid test instruments to monitor the sign language acquisition of Deaf children in different countries, very few tests offering strong evidence for their psychometric properties are commercially available. A German Sign Language (DGS) test that focuses on linguistic structures acquired in preschool- and school-aged children (4–8 years old) is urgently needed. The present study uses as a template a test which has sound psychometric properties and has been standardized on another sign language as a starting point for tests of sign languages that are less documented, such as DGS.

This book makes a novel contribution to the field by examining linguistic, cultural, methodological, and theoretical issues in the process of the adaptation from the source language test to the target language test, and by providing a model for future test adaptations. It also includes concrete steps for the test development and adaptation process.

Adaptation and Evaluation of a German Sign Language Test addresses students and researchers alike who are involved in sign language test development and adaptation. It also provides a comprehensive summary in German.

STAATS- UND UNIVERSITÄTS-
BIBLIOTHEK
HAMBURG
CARL VON OSSIETZKY

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg
ISBN 978-3-937816-79-1