Approaching Sign Language Test Construction: Adaptation of the German Sign Language Receptive Skills Test

Tobias Haug*

University of Applied Sciences of Special Needs Education Zurich (HfH)

Received August 24, 2010; revisions received November 16, 2010; accepted November 28, 2010

There is a current need for reliable and valid test instruments in different countries in order to monitor deaf children's sign language acquisition. However, very few tests are commercially available that offer strong evidence for their psychometric properties. A German Sign Language (DGS) test focusing on linguistic structures that are acquired in preschool- and school-aged children (4-8 years old) is urgently needed. Using the British Sign Language Receptive Skills Test, that has been standardized and has sound psychometric properties, as a template for adaptation thus provides a starting point for tests of a sign language that is less documented, such as DGS. This article makes a novel contribution to the field by examining linguistic, cultural, and methodological issues in the process of adapting a test from the source language to the target language. The adapted DGS test has sound psychometric properties and provides the basis for revision prior to standardization.

Introduction

Internationally, the education of deaf children has changed over the past decades with the emergence of bilingual and bicultural programs in the United States (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, in most cases as a means upon which to build their knowledge of the written (and spoken) forms of the majority language as a second language (L2).

Research has been conducted in several countries to evaluate the effectiveness of such programs and/or to investigate the relationship between a sign language as a first language (L1) and literacy skills of the majority language as L2 (e.g., United States: Hoffmeister, 2000; Strong & Prinz, 1997, 2000; Germany: Mann, 2006; Switzerland: Niederberger, 2004, 2008). The results suggest a positive correlation between sign language skills and written skills in the majority language. However, Plaza-Pust and Morales-López (2008) draw attention to some of the shortcomings of existing bilingual programs, namely, "the 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 this article is the fact that a range of *evaluation procedures* are needed for bilingual programs, including tests to evaluate sign 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; United Kingdom: Herman, 1998; United States: Mann & Prinz, 2006).

Current Bilingual Deaf Education in Germany

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). Since then, bilingual methods

^{*}Correspondence should be sent to Tobias Haug, Sign Language Interpreter Training Program, University of Applied Sciences of Special Needs Education Zurich, Schaffhauserstrasse 239, POB 5058, CH-8050 Zurich, Switzerland (e-mail: tobias.haug@hfh.ch).

[©] The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

have become more accepted in Germany but still constitute a minority within the educational approaches for deaf children currently used 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, 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 a hearing loss that are mainstreamed is about 20%, and 90% of those are hard of hearing (Günther et al., 2009). As to the modes of communication used in German schools for the deaf (Große, 2003), although the majority of deaf institutions (90%) have the mastery of spoken language as their primary goal, in about 60% of the classes manual means of communication are included to some extent. 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 Signed German (LBG) to German Sign Language (DGS: Deutsche Gebärdensprache), is used in their institution (Haug & Hintermair, 2003).

Sign Language Evaluation

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 United States use informal descriptive evaluations of 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 respondents (from nine institutions) reported that sign language skills are evaluated in their institution on a regular basis. As in the United States, the absence of any standardized measure of sign language means that evaluation procedures are mostly informal, such as observations in class or video analysis (Haug & Hintermair, 2003). As testing and monitoring the DGS development of deaf children, particularly in the early ages, is of great importance, there is a clear need for a sign language test that measures a range of linguistic devices in DGS that are important for language acquisition from the age of 3 onward.

Considering the state of research in this field, test adaptation is a practical approach that 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 that has been standardized and has sound psychometric properties as a template for adaptation thus provides a starting point for tests of sign languages that are less well 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.

Within the frame of this study, the standardized British Sign Language (BSL) Receptive Skills Test (Herman, Holmes, & Woll, 1999) was used as a template for adaptation to DGS. The goal was to establish sound psychometric properties (item and distractor analysis, homogeneity index, reliability, evidence of relations to an external variable, content validity) and investigate the relationships between the deaf children's raw scores and other variables (gender, age of sign language exposure, parental hearing status, chronological age) that are important for test development and adaptation. The results should be used as a basis for a standardization of the DGS Receptive Skills Test.

Adapting Sign Language Tests

Hambleton (1994, 2005) and Hambleton and Patsula (1998) define *adaptation* as the entire process beginning with the source test (i.e., BSL test) and ending with the target test (i.e., DGS 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.

Only one publication directly addresses the issue of test adaptation from a source to a target sign language (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: (a) language-specific issues and (b) culture-related issues. In relation to languagespecific issues in the adaptation of the Test Battery for American Sign Language (ASL) Morphology and Syntax (Supalla et al., 1995, unpublished) to Australian Sign Language (Auslan; Schembri et al., 2002), morphosyntactic differences between the two languages 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, whereas 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), 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 that 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. Whereas 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. The available literature suggests that these differences between sign languages are rather lexical than morphological because it deals only with the number of available negator signs in the different sign languages.

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 (Prinz, Strong, & Kuntze, 1994) with their adapted versions in Swiss French Sign Language (Niederberger, 2004). Prinz et al. (2005) report results in relation to participants' responses for one of the items from the story comprehension task concerned with obtaining a driver's license. Whereas most American participants showed no difficulties with this item, it was reported to be one of the harder items for Swiss French participants. The researchers hypothesized that this divergence may be because of the different significance of having a car in the two cultures.

Review of Sign Language Tests

A number of sign language tests have been developed with the goal of evaluating sign language development of deaf children across different age ranges in different sign languages. But only very few of these tests have been published. The tests that have not yet been published are only available directly from the authors.

An extensive review of the available sign language tests served as a basis for the decision on which test would be used for adaptation to DGS (Anderson & Reilly, 2002; Baker & Jansma, 2005; Bizer & Karl, 2002; Fehrmann, Huber, Jäger, Sieprath, & Werth, 1995a, 1995b; Herman et al., 1999, 2004; Hermans, Knoors, & Verhoeven, 2010; Hoiting, 2009; Huber, Sieprath, & Werth, 2000; Johnston, 2004; Maller, Singleton, Supalla, & Wix, 1999; Mounty, 1993, 1994; Woolfe, Herman, Roy, & Woll, 2010). Three important criteria needed to be met: (a) reported psychometric properties, (b) testing of the development of language comprehension—an important but often neglected area in language testing, and (c) focusing on an age group where standardized testing formats can be used (>3 years). Only the BSL Receptive Skills Test met these criteria.

The Template: The BSL Receptive Skills Test

The BSL Receptive Skills Test (Herman et al., 1999) is designed for children aged 3–11 years. The BSL Receptive Skills Test focuses on selected aspects of morphology and syntax of BSL, and it is not based on a test for spoken English. It consists of a vocabulary check and a video-based receptive skills test.

Vocabulary check: The children confirm their knowledge of the 22 vocabulary items used in the main test through a simple picture-naming task.

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 United Kingdom. The items of this test evaluate children's receptive knowledge of a variety of BSL syntactic and morphological structures: (a) spatial verb morphology (e.g., agreement verbs and spatial verbs with whole entity classifiers), (b) number and distribution, (c) negation, (d) size and shape specifiers (SASS), (e) noun-verb distinction, and (f) handling classifiers.

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 was 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 participants 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).

Sign Language Acquisition

Previous studies of sign language acquisition provided an important basis for the adaptation of the DGS Receptive Skills Tests. For DGS acquisition, there is only one study available, which concerns the acquisition of verb agreement by two deaf children aged 2.2-3.4 years (Hänel, 2003, 2005). Therefore, studies on the acquisition of other sign languages were reviewed in order to provide an overview of the emergence and *mastery* of the linguistic structures represented in the BSL test. The main focus was on studies that covered the age range of 4-8 years, the age group of the adapted DGS test. These studies referred to ASL (Anderson & Reilly, 1997, 2002; Bellugi, van Hoek, Lillo-Martin, & O'Grady, 1988; Hoffmeister, 1992; Martin & Sera, 2006; Reilly, 2006; Reilly & Anderson, 2002; Schick, 1987, 1990; Slobin et al., 2003), Auslan (de Beuzeville, 2004, 2006), BSL (Morgan, Barrière, & Woll, 2003, 2006; Morgan, Herman, Barrière, & Woll, 2008; Morgan & Woll, 2002, 2003), Sign Language of the Netherlands (NGT; Slobin et al., 2003), Brazilian (Bernardino, 2005), and LIS (Pizzuto, 2002).

Together these studies provided an overview of the emergence and mastery of the targeted linguistic structures that should be represented in the adapted DGS test. Therefore, it is argued that the findings from other sign languages can be used as a basis for making informed decisions about what should be represented in the adapted DGS test.

Cross-linguistic Differences

In the next step, studies on the linguistic structures in DGS that are represented in the BSL test were analyzed in order to look for similarities and differences across the two sign languages (Becker, 2003; Glück, 2001, 2005; Glück & Pfau, 1997a, 1997b, 1998; Happ,

2005; Happ & Vorköper, 2005; Papaspyrou, von Meyenn, Matthaei, & Herrmann, 2008; Perniss, 2001, 2007; Pfau, 2001, 2004; Pfau & Quer, 2002, 2007; Pfau & Steinbach, 2005, 2006; Rathmann, 2003; Rathmann & Mathur, 2002).

For example, the review of the literature on spatial verb morphology in DGS suggests that it is similar to BSL in the linguistic features it has at its disposal. The studies on DGS suggest that many of the structures described in other sign languages are also available in DGS (e.g., SASS and handling classifiers), whereas others probably do not exist in DGS (derivationally related noun-verb pairs; Becker, 2003). In turn, other structures in DGS not only have some features in common with BSL, such as verb agreement, but also have language-specific features, such as the Person Agreement Marker (PAM) AUF (Rathmann, 2003; Rathmann & Mathur, 2002). 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 PAM AUF is used (Rathmann, 2003; Rathmann & Mathur, 2002).

Adaptation Process

There were several stages to the adaptation process, which are described below.

Review and Revision of Test Materials

The picture materials—for both the vocabulary check and 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). These changes were mostly culture related, for example, the steering wheel of a British car needed to be moved from the right to the left side of the car and a British round red mailbox needed to be replaced by a square yellow German mailbox.

Pilot 1

The objective of Pilot 1 was to check for regional variations of DGS lexical items (vocabulary items). Thirteen informants in three regions (where the test-ing later took place) were included (age range = 12-57

years; M = 31; 4 males, 9 females). Most of the items from the vocabulary check, which depict simple nouns, showed no regional variation. There was some variation in the signs JUNGE (boy), KIND (child), HUND (dog), MUTTER (mother), and TEDDY-BÄR (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.

Adaptation of Items

The first version of the adapted computer-based 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. The 10 additional items were developed because of the possibility that some of the items of the original BSL test do not work in the same way for DGS (i.e., order of acquisition is different) or that 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). During the development of the 10 additional items, the main goal was to cover the same linguistic structures (e.g., negation, spatial verb morphology) covered in the BSL test. Thus, the new items paralleled those in the original test: for example, items on spatial morphology and with varying levels of difficulty. The item order followed the order of the BSL items followed by the newly developed DGS items.

Filming of Test

The test instructions and the test items were filmed. With respect to lexical variation, because the collected data of Pilot 1 did not reveal any clear regional variation and the creation of three or four different test versions would have been complicated, it was decided in collaboration with the deaf signer who modeled the test materials to use *one* single variant throughout the entire test. The variant was decided by the deaf person. If children did not know these signs/variants in the vocabulary check, they were 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 he/she knew and understood the sign.

Programming Test Interface

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 (a) general introduction and test instructions, followed by the vocabulary check, where the test administrator had to mark the vocabulary on a checklist, which was adapted from the BSL test; (b) a training session, that is, the deaf signer on the video "taught" the children the lexical signs used in the DGS Receptive Skills Test for which regional variants were identified during Pilot 1; and (c) the receptive skills test, which had an introduction followed by three practice items and then the 50 test items.

Pilot 2

Deaf adults were involved in this pilot as a reference for adult users of DGS. 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. A total of five deaf adults were tested. 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 were tested individually. Notes were taken during the testing in order to record feedback from the informants. Because the test results are stored automatically, specific focus was placed on gathering feedback about informants' views of the pictures and signed stimulus sentences. The informants were also asked to explain the reasons for their response choice. At the end, both the feedback (qualitative) and the test data (quantitative) were analyzed.

In addition, a pilot with *non-signing hearing children* was conducted. The objective of this pilot was to make observations on the user-friendliness of the test interface, for example, was the test easy to navigate or were there any general problems in the structure of the test. A total of 13 children were tested, ranging in age from 4.8 to 7.10 (M = 5.8).

Revision of First Test Version

Based on the results of Pilot 2 with deaf adults, the following revisions were made¹:

- (1) Re-filming of 10 items
- (2) Revisions of 9 pictures

Observations made during Pilot 2 testing of the non-signing hearing children resulted in various changes to enhance the user-friendliness of the test for the target group, for example, simplifying video navigation by making the buttons larger and changing the interface.

This revised version consisted of 49 items, including 3 practice items. Because the linguistic status of derivationally related noun-verb pairs in DGS is undetermined (Becker, 2003), these four items were removed.

Planning of Main Study

Five schools participated in this study. Once the school administration and the teachers had agreed that they wanted to take part in this study, an application pack consisting of information sheets, background questionnaires (including the teachers' rating of their own DGS skills and the rating of the deaf students' receptive and productive DGS skills), and consent forms 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. 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 on the deaf child's preferred languages. Two questionnaires were completed by the teachers: one for each individual child and a second one requesting general information about the school.

Parents' hearing status	Male participants (n)	Female participants (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)

Table 1Description of the sample (N = 54)

Main Study

Administering the test took approximately 30 min for the younger children (3.9-5.6) and 20 min for the older children (>5.7). All test instructions were provided in video format on the computer-based test which was displayed on a laptop computer. At the end of each test session, the results were saved to an individual folder on the computer's hard drive. During the session, a scoring sheet for the vocabulary test was used to check the child's familiarity with the vocabulary.

Statistical Assumptions

A test was used in order to determine the normal distribution of the sample in order to apply parametric or nonparametric statistical testing methods (Kiess, 1996). Using a histogram of the variable raw score with a normal curve overlaid, the results revealed that the sample is left skewed and thus does not represent a normally distributed sample (M = 30.72, SD = 10.15, N = 54). Therefore, nonparametric statistical procedures were applied.

It was decided to use an alpha level of .05 (twotailed) 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 (a) .10 as small, (b) .30 as medium, and (c) .50 as large. The statistical package SPSS (Statistical Package for the Social Sciences) was used for the analysis (e.g., Gaur & Gaur, 2006).

Results of the DGS Test Adaptation

The Sample

A total of 54 deaf children were tested. Thirty-four (63%) came from deaf families with at least one deaf parent, and 20 (37%) came from hearing families. The

whole sample consisted of 29 boys and 25 girls between 3.9 and 10.10 years of age (M = 7.0), who attended one of five schools or kindergarten programs that (a) implemented a bilingual philosophy using DGS as the language of instruction, (b) implemented a bilingual pilot classroom with subsequent use of DGS in other classes across the school, or (c) used "signing" to a certain degree as the means of instruction, ranging from DGS to manual communication, such as LBG. Table 1 provides a descriptive overview of the entire sample.

Reported hearing losses for these 54 children were (a) 1 child with a mild hearing loss (25–40 dB), (b) 2 with a moderate hearing loss (40–70 dB), (c) 29 with a severe hearing loss (70–100 dB), and (d) 18 with a profound loss (>100 dB). No information was provided for four children.

For the research questions regarding quality of the test instrument (e.g., item and distractor analysis, homogeneity index, reliability), only the subgroup of deaf children of deaf parents was 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, addressing the test performance of the deaf children, was analyzed with data from the whole sample or by comparison of both subgroups. Content validity will be investigated in the Discussion section.

The sociodemographic information used in this study is based on the questionnaires completed by the teachers, as introduced in the Adaptation Process

DGS (<i>n</i>)	Other sign languages (<i>n</i>)	German (<i>n</i>)	LBG (<i>n</i>)	Home signs (<i>n</i>)	Two Slavic languages (<i>n</i>)	Three other European spoken languages (<i>n</i>)	Three other non-European/ Slavic spoken languages (<i>n</i>)
34	5	31	19	11	3	8	3

Table 2 Languages used in the children's home (multiple responses are not shown; N = 54)

Note. DGS, German Sign Language; LBG, Signed German.

section. In Table 2, an overview of all languages and means of communication used in the children's home is presented.

Results of Item Analysis

The item analysis consists of the item facility value p_i and discrimination coefficient r_{ii} . A total of 49 items were analyzed. The results of the item analysis contributed to whether items were to be removed from the item pool for subsequent analysis (or suggested for revision with subsequent new piloting for a standardization study).

Items were retained in the item pool when they met the following criteria (Fisseni, 2004; Lienert & Raatz, 1998):

(1) Items with a facility value p_i between .25 and .90 (2) Items with an item discrimination coefficient $r_{it} > .25$

A total of 10 items were removed from the item pool for subsequent analysis as they did not meet these criteria. Some items might be considered for revision, requiring new piloting. The reason for revision is content driven because some of the items represent important linguistic structures that are relevant to be included in a DGS test that targets deaf children's comprehension of morphosyntactic structures from 4 years onward.

Fit of the Newly Developed Items

Ten additional items were developed in the course of the adaptation. One item had been removed from the item pool after the Pilot 2 study because it represented a linguistic structure (derivationally related noun–verb pairs) that probably does not exist in DGS. Of these nine items, four were suggested for removal (or revision) based on item analysis.

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_{ii} . The results of the distractor analysis contributed to deciding 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). All distractors, even those for items that would later be removed from the item pool based on the item analysis, were analyzed.

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. Some distractors need to be revised prior to a standardization of the DGS test.

Homogeneity of the DGS 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; instead, 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 index H was investigated by applying an inter-item correlation (Bortz & Döring, 2005; Fisseni, 2004), using the Pearson product-moment correlation. 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.

Evidence for Reliability

The internal consistency of the test was calculated by the statistical analysis of Cronbach's alpha (Rust & Golombok, 2000). 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.

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 in one of the background questionnaires. These ratings were available for 31 of the 34 deaf children of deaf parents.

The results reveal that there is a statistically significant positive 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 on the adapted DGS test is correlated with better receptive DGS skills as rated by teachers. 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 because receptive skills are the goal of the adapted DGS test.

These correlations should be treated with caution because 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. They were provided with a scale describing five different skill levels (minimum score 1, maximum score 5).

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 because the latter are more likely to use DGS as their preferred language.

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).

Test Performance of Deaf Children

This section addresses the issue of whether the test performance (for all items) is influenced by (a) gender of the children, (b) age of sign language exposure, (c) parental hearing status, and (d) chronological age.

Fisher's exact test (Table 3) was applied to see if there is a significant relationship between the raw score and (a) gender of children, (b) age of sign language exposure, (c) parents' hearing status, and (d) chronological age.

Results indicate a nonsignificant relationship between raw scores and gender, which was confirmed by a Mann–Whitney U test (U = 284.5, p = .175). In contrast, a significant relationship between raw scores and the three measures of age of sign language exposure, parents' hearing status, and chronological age ($\alpha = .05$) was found, and the nature of this relationship was investigated further.

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

	Gender of children ($n = 54$)	Age of sign language exposure $(n = 35)$	Parents' hearing status ($n = 54$)	Chronological age $(n = 54)$
Raw score (in categories)	p = .090	$p < .001^{a}$	$p = .011^{a}$	$p = .009^{a}$
a				

Table 3 Fisher's exact test across raw score and gender, age of sign language exposure, parents' hearing status, and chronological age

^aExact significance, two-tailed.

the situation is different because only 5% have deaf parents (Mitchell & Karchmer, 2004) and, most likely, 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 an 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, Lock, & Kazmi, 2002). Therefore, the degree to which early L1 acquisition of DGS versus 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. 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 eight children, all with hearing parents, whose mean age was 6.5 (range = 5.2-6.5).

A univariate analysis of variance (ANOVA) was computed to compare the difference in test performance between the early and late exposure group.² The results reveal that the children with early exposure performed significantly better (M = 36.04) than the late exposure group (M = 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 significantly different (F = 3.11, df = 1, p = .087).

In a next step, a univariate analysis of covariance (ANCOVA) was calculated with chronological age as a control variable (covariate) in order to investigate whether chronological age also accounts for performance differences on the adapted DGS test. Controlling for chronological age, the main factor early versus late exposure explains more of the performance differences between the two groups (F = 23.42, df = 1, p < .001) than the covariate chronological age (F = 8.4, df = 1, p = .007). Chronological age still has an influence on performance but not as strong as age of exposure.³

In sum, the results suggest that early exposure to DGS has an impact on test performance. However, the results only explain a relation, not causation. The control variable chronological age also has an impact on test performance but not as strong as the factor age of exposure. The parents' hearing status has an overlap with the two groups of early and late exposure to DGS, and therefore, the variable age of exposure 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.

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

Examination of the learning trajectories of children of deaf and hearing parents provided a general descriptive overview of both subgroups. The deaf children of hearing parents (n = 20) 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 (n = 34) 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. 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 complete 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, it is not entirely certain whether the learning trajectories of deaf children of deaf parents and those of hearing parents vary empirically. This issue will be further expanded in the Discussion section.

The Mann–Whitney U test was applied to investigate the difference in test performance between children with deaf and 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, mean 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, mean age = 7.4). The different mean ages (6.10 vs. 7.4) in both groups are not significantly different (U = 268, p =.197). Thus, there is evidence that parentage has a statistically significant impact on test performance, but it is not clear what leads to better test performance.

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 issue concerning a correlation between chronological age and raw scores, both variables were correlated, for the whole sample, then separately for the subgroups of children with deaf parents and hearing parents. 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.

For the subgroup of deaf children of deaf parents, the Spearman rank correlation coefficient indicates a strong effect size of the correlation ($r_{\rm s} = .681$, 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.

In a next step, the deaf children with deaf and hearing parents together were grouped into three age bands and then correlated with raw score. In the first two bands (3.9-5.11 and 6.0-7.11), the correlation between chronological age and raw scores is strong (first band: $r_s = .532$, p = .023; second band: $r_s =$.593, p = .005). 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 ($r_s = .394$, p = .131), suggesting that from 8 years onwards there is no relation between chronological age and raw scores, and 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 (eight children in each group).

These results provide important evidence regarding the issue of whether age correlates with the test scores.

Discussion

Effectiveness of Items

The findings of the empirically driven research questions provide a basis for item selection for a standardization study (i.e., item and distractor analysis, homogeneity index, and Cronbach's alpha). As for the distractor analysis, the results indicated that the distractors for some items would need to be revised before a standardization. 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) because the linguistic structures representing spatial concepts in front, left-right in whole entity classifiers, are acquired rather late (>10 years old) and have most likely therefore not been solved by many children of this sample (see also Means to Differentiate Among Participants section). These items should instead 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.

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. This issue should be addressed prior to the standardization of this DGS test.

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, it 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.

Other studies have also addressed the issue of whether teachers' ratings of deaf children's sign language competence provide 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. Herman and Roy 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 children's test scores did not match with the impressions of teachers based on their everyday interaction with the children. All deaf and hearing teachers seemed to have good Auslan skills (Johnston, 2004). 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 and also for the standardization of the DGS test, it would be advisable to revise and standardize the selfrating scale used in this article in order to get a better measure of the teachers' DGS skills to include it as an external source of validation-in addition to the standardized DGS vocabulary test Perlesko (Bizer & Karl, 2002), which was not available when this study was conducted.

Content Validity

Content validity is defined as being present if, for example, the test items (and the test a whole) represent the linguistic structures to be tested (Davies et al., 1999). This issue was approached by reviewing (a) studies on sign language acquisition and (b) studies on DGS structures represented in the BSL template (e.g., negation, spatial verb morphology). 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.

Other Variables Explaining Performance Differences

Other variables that are important for successful test adaptation involve factors that have been identified as potentially affecting scores (Herman & Roy, 2006; Johnston, 2004). These variables include (a) gender, (b) age of exposure, (c) parents' hearing status, and (d) chronological age.

 The adapted DGS test does not show any gender differences in the performance of male and female participants. (2) The variable age of exposure to DGS, represented by an early exposure group (0-3 years old) and a late 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 length of use of a sign language (i.e., signing age).

In studies of the impact of early L1 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 L1 acquisition (e.g., Mayberry et al., 2002). The length of DGS use could not be investigated because of the limited information available. The variable signing age 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 (i.e., deaf children of deaf parents outperforming deaf children of hearing parents), but because it was not possible to investigate the effect of signing age, the source of the difference is not clear (e.g., early input).

The deaf children of deaf parents also reach their highest scores when they are between 6 and 7 years old, suggesting that the adapted DGS test is not sensitive enough for children >7 years old. This suggests that more difficult items should be developed and piloted prior to standardization.

(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 ASL (Enns & Zimmer, 2009). This is in contrast to the BSL test, which is standardized and differentiates between children from 3 to 11 years old (Herman et al., 1999).

The Reference and the Target Groups of Sign 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 to 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 be able to comprise as "homogeneous" and as large a group as possible for the standardization study. One could argue that only deaf children (and may be 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.

This issue will later be linked to the estimated size of the norming sample for the standardization study.

Means of Differentiation among 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 normreferenced 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 that 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.

(a) Item complexity: Items representing spatial concepts that have been identified as needing revision or new sampling could be used in a future test version as a means of differentiating between younger and older children. These seven items represent different spatial concepts such as *in front*, *behind*, *top-right*, *belom-left*, or *inside-left*, which are acquired relatively late (>10 years; Morgan et al., 2008; Slobin et al., 2003). These items

were not comprehended by many children in this article; 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. 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.

(b) High and low frequency: Relatively low-frequency structures in a language also offer a means of differentiating 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 high- and low-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 (e.g., Morgan et al., 2008).

Defining the Norming Sample

Estimating an exact number required for a norming sample of the adapted DGS test is hard because too many variables would need to be defined in determining what constitutes the entire "population of hearingimpaired children in Germany" (e.g., deaf, hard-ofhearing 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 10,000-12,700 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 because 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 other empirical studies (Herman et al., 1998; Hermans et al., 2010): (a) 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 (b) 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, that is, at least 180 children. This constitutes the minimum of potential participants to conduct a standardization, but it should be attempted with more children if they are available. Yearly intervals in the younger age (till age 6) are important because language development is more marked at these ages (Herman et al., 1998; Hermans et al., 2010). Herman et al. 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 longterm goal.

Future Research

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

(a) 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. (b) To be able to develop and adapt tests for DGS, more acquisition studies of DGS comprehension and production are needed. (c) More cross-linguistic sign language research is needed in order to get a clearer understanding of the differences and similarities between sign languages and their acquisition. (d) Also of importance is more research on variability and acceptability of linguistic structures beyond the lexical level, that is, too little is known about morphosyntactic structures and how they are used at the sentence and discourse levels.

Conclusion

Among the main contributions of this article are new insights into the cultural, linguistic, and methodological considerations necessary for future sign language test adaptation. The interconnected cultural, linguistic, and methodological issues were addressed at different stages of test adaptation. 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. Sign language test adaptation is not a "quick and dirty" approach to designing a test: the result can produce a valid and reliable instrument to be used in schools.

The increasing number of sign language acquisition studies can serve to inform test adaptation and development, but 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 thus reciprocal for sign language acquisition and test development and will contribute to further development in this field.

Notes

1. 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-skillstest.html.

2. At first a nonparametric test for between-participant 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 raw score as the dependent variable 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 whether the variable chronological age also explains 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.

3. Signing age, that is, the length of use/exposure of a sign language could not be investigated because of the empirical data available of this study. Signing age has been included in studies of the impact of early L1 acquisition on language processing by deaf adults as a control variable (e.g., Mayberry et al., 2002).

Conflict of Interest

No conflicts of interest were reported.

References

- 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, 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. Hörgeschädigtenpädagogik, 61, 15–20.
- Baker, A., & Jansma, S. (2005). NGT-OP Een Screening Instrument voor Peuters in de Nederlandse Gebarentaal em Handleiding. Amsterdam, The Netherlands: University of Amsterdam Internal Publication.
- Becker, C. (2003). Verfahren der Lexikonerweiterung in der Deutschen Gebärdensprache. 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). Graduate School of Arts and Sciences, Boston University (ProQuest UMI number: 3186488).
- Bizer, S., & Karl, A.-K. (2002). Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schriftund Lautsprache. (Unpublished doctoral dissertation). School of Education, Hamburg University, Germany.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage). Heidelberg, Germany: Springer.
- Bortz, J., & Döring, N. (2005). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (3. Auflage). Heidelberg, Germany: Springer.
- 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, 106–148.
- 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.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- 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, 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.
- Enns, C., & Zimmer, K. (2009). Research study: Adapting the British Sign Language Receptive Skills Test into American Sign Language. (Summary Report). Retrieved 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). (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). (Unpublished manuscript). RWTH-Aachen, Germanistisches Institut & Neurologische Klinik, Projekt DESIRE.
- Fisseni, H.-J. (2004). Lehrbuch der psychologischen Diagnostik-Mit Hinweisen zur Intervention (3. Auflage). Göttingen, Germany: Hogrefe Verlag.

- 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.
- Glück, S. (2001). Morphosyntaktische Eigenschaften der Klassifikation in Deutscher Gebärdensprache. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000—Theorie* und Anwendung (pp. 127–145). Hamburg, Germany: Signum Verlag.
- Glück, S. (2005). Dimensionen—Klassifikation und Spezifizierung in Gebärdensprachen und Lautsprachen. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* (pp. 179–200). Hamburg, Germany: Helmut Buske Verlag.
- Glück, S., & Pfau, R. (1997a). Eine Klasse für sich: Klassifizierende Verben in Deutscher Gebärdensprache. Zeitschrift für Sprachwissenschaften, 16, 181–208.
- Glück, S., & Pfau, R. (1997b). Einige Aspekte der Morphologie und Morphosyntax in Deutscher Gebärdensprache. 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.
- Große, K.-D. (2003). Das Bildungswesen für Hörbehinderte in der Bundesrepublik Deutschland—Daten und Fakten zu Realitäten und Erfordernissen. Heidelberg, Germany: Universitätsverlag Winter GmbH.
- Günther, K. B. (1999). Bilingualer Unterricht mit gehörlosen Grundschülern—Zwischenberich zum Hamburger Bilingualen Schulversuch. Hamburg, Germany: Verlag hörgeschädigter Kinder.
- Günther, K. B., & Hennies, J. Bilingualer Unterricht in Gebärden-, Laut- und Schriftsprache mit gehörlosen SchülerInnen in der Primarstufe: Zwischenbericht zum Berliner Bilingualen Schulversuch. Berlin, Germany: 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. Hamburg, Germany: Signum Verlag.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal* of Psychological Assessment, 10, 229–244.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17, 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 crosscultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.

- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45, 153–171.
- Hänel, B. (2003). Der Erwerb der Deutschen Gebärdensprache als Erstsprache – Die frühkindliche Sprachentwicklung von Subjektund Objektverbkongruenz in DGS. (Unpublished doctoral dissertation). Department of Linguistics, Hamburg University, Germany.
- 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* (pp. 201–232). Hamburg, Germany: Helmut Buske Verlag.
- Happ, D. (2005). Manuelle und nicht manuelle Module der Deutschen Gebärdensprache (DGS): Linguistische. In H. Leuninger & D. Happ (Eds.), Gebärdensprachen: Struktur, Erwerb, Verwendung (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. In H. Leuninger & D. Happ (Eds.), Gebärdensprachen: Struktur, Erwerb, Vermendung (pp. 87–110). Hamburg, Germany: Helmut Buske Verlag.
- Haug, T., & Hintermair, M. (2003). Ermittlung des Bedarfs von Gebärdensprachtests für gehörlose Kinder—Ergebnisse einer Pilotstudie. Das Zeichen, 64, 220–229.
- 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, 138–147.
- 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, 3–8.
- Herman, R. (2002). Assessment of BSL development. (Unpublished doctoral dissertation). City University London.
- Herman, R., Grove, N., Holmes, S., Morgan, G., Sutherland, H., & Woll, B. (2004). Assessing BSL development: Production test (narrative skills). London, UK: City University Publication.
- 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., & Roy, P. (2006). Evidence from the wider use of the BSL Receptive Skills Test. *Deafness and Education International*, 8, 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, 107–119.
- 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. (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.
- Hoiting, N. (2009). The acquisition of sign language by deaf toddlers (Doctoral dissertation). University of Groningen, the Netherlands.
- Huber, W., Sieprath, H., & Werth, I. (2000). Sprachdiagnostik in der Deutschen Gebärdensprache. In H. Leuninger & K. Wempe (Eds.), *Gebärdensprachlinguistik 2000—Theorie und Anwendung* (pp. 255–276). Hamburg, Germany: Signum Verlag.
- 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, 57–81.
- Kiess, H. O. (1996). Statistical concepts for the behavioral sciences (2nd ed.). Boston, MA: Allyn & Bacon.
- Krausneker, V. (2004). Viele Blumen schreibt man "Blümer"— Soziolinguistische Aspekte des bilingualen Wiener Grundschul-Modells mit Österrreichischer Gebärdensprache und Deutsch. Hamburg, Germany: Signum Verlag.
- Lewis, W. (1995). Bilingual teaching of deaf children in Denmark— Description of a project 1982–1992. Aalborg, Denmark: Døveskolernes Materialcenter.
- Lienert, G., & Raatz, U. (1998). Testaufbau und Testanalyse (6. Auflage). Weinheim, Germany: Beltz Psychologie Verlags Union.
- Mahshie, S. N. (1995). Educating deaf children bilingually—With insights and applications from Sweden and Denmark. Washington, DC: Gallaudet University Press.
- 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, 249–269.
- Mann, W. (2006). Examining German deaf children's understanding of referential distinction in written German and German Sign Language. (Unpublished doctoral dissertation), San Francisco State University & University of California, Berkeley.
- Mann, W., & Prinz, P. (2006). The perception of sign language assessment by professionals in deaf education. *American Annals of the Deaf*, 151, 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, The Netherlands: John Benjamins Publishing Company.
- 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, 391–402.
- Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Development: Linguistic ability and early language exposure. *Nature*, 417, 38.
- 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, 138–163.

- 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, 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., & Woll, B. (2002). The development of complex sentences in British Sign Language. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 255–276). Amsterdam, The Netherlands: 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.
- Newport, E. L. (2002). Critical periods in language development. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 737–740). London, UK: Macmillan.
- Niederberger, N. (2004). Capacités langagières en Langue des Signes Française et en français écrit chez l'enfant sourd bilingue: quelles relations? (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. Morlas-López (Eds.), Sign bilingualism—Language development, interaction, and maintenance in sign language contact situations (pp. 29–50). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- 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, Gallaudet University, Washington, DC, March 2005.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Papaspyrou, C., von Meyenn, A., Matthaei, M., & Herrmann, B. (2008). Grammatik der Deutschen Gebärdensprache aus der Sicht gehörloser Fachleute. Hamburg, Germany: Signum Verlag.
- Perniss, P. (2001). Numerus und Quantifikation in der Deutschen Gebärdensprache (Unpublished master thesis). Germany: University of Köln.
- Perniss, P. (2007). Achieving spatial coherence in German Sign Language narratives: The use of classifiers and perspective. *Lingua*, 117, 1315–1338.
- Pfau, R. (2001). Typologische und strukturelle Aspekte der Negation in Deutscher Gebärdensprache. In H. Leuninger

& K. Wempe (Eds.), *Gebärdensprachlinguistik 2000—Theorie* und Anwendung (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., & Steinbach, M. (2005). Plural formation in German Sign Language: Constraints and strategies. In H. Leuninger & D. Happ (Eds.), *Gebärdensprachen: Struktur, Erwerb, Verwendung* (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.
- 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, The Netherlands: 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. Morlas-López (Eds.), Sign bilingualism—Language development, interaction, and maintenance in sign language contact situations (pp. 333–380). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- 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, CA: 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). Oxford: Oxford University Press.

- Reilly, J., & Anderson, D. (2002). FACES: The acquisition of nonmanual morphology in ASL. In G. Morgan & B. Woll (Eds.), *Directions in sign language acquisition* (pp. 159–182). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Rust, J., & Golombok, S. (2000). Modern psychometrics—The science of psychological assessment (2nd ed.). London: Routledge.
- 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, 18–40.
- Schick, B. (1987). The acquisition of classifier predicates in American Sign Language. (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN (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.
- 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.
- 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.
- Strong, M., & Prinz, P. (1997). A study of the relationship between ASL and English literacy. *Journal of Deaf Studies* and Deaf Education, 2, 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., Newport, E., Singleton, J., Supalla, S., Coulter, G., & Metlay, D. (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), Department of Psychology, University of Trieste, Italy.
- Tomasello, M. (2003). Constructing a language—A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.
- 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, 322–331.