Exploring Signed Language Assessment Tools in Europe and North America

Charlotte Enns
Faculty of Education
University of Manitoba
Winnipeg, Manitoba R3T 2N2
CANADA
charlotte.enns@umanitoba.ca

Tobias Haug
Dept. of Pedagogical and Therapeutic Professions
University of Applied Sciences of
Special Needs Education
Zurich, SWITZERLAND
tobias.haug@hfh.ch

Rosalind Herman
Division of Language &
Communication Science
City University
London, UNITED KINGDOM
R.C. Herman@city.ac.uk

Robert Hoffmeister
Center for the Study of Communication & the Deaf
Boston University
Boston, Massachusetts
UNITED STATES
rhoff@bu.edu

Wolfgang Mann
Department of Education
University of Roehampton
London, UNITED KINGDOM
Wolfgang.Mann@roehampton.ac.uk

Lynn McQuarrie
David Peikoff Chair of Deaf Studies
Director, Western Canadian Centre for Deaf Studies
University of Alberta
Edmonton, Alberta T6G 2G5 CANADA
lynn.mcquarrie@ualberta.ca

Acknowledgements:

Test development requires teamwork, and the work we report on in this chapter would not have been possible without significant assistance from our colleagues, graduate students, research assistants, and technicians, as well as the support we received from schools, parents and children to complete the testing. In addition, we would like to acknowledge the funding support we have received to conduct our research studies in signed language assessment.

Abstract:

The development of valid and reliable assessment tools to measure the acquisition of natural signed languages is of practical as well as theoretical significance. This chapter describes a selection of tests that are currently available or under development to assess several signed languages (American Sign Language, British Sign Language, German Sign Language, and Swiss German Sign Language) at the phonological, lexical, grammatical, and discourse levels. In addition to test descriptions we discuss issues pertaining to test development procedures, test formats, normative sample composition, and the use of web-based technology. The importance of assessment in guiding instruction in deaf education is emphasized throughout.

Keywords: assessment, signed language acquisition, phonology, morphology, syntax, target population, reliability, validity, web-based technology

Over the past several years significant progress has been made towards developing tests to assess the acquisition of natural signed languages. As many of these tests are still works in progress, the need for valid and reliable assessment tools of signed languages for both practical and theoretical purposes continues to be important. The purpose of this chapter is to provide state-of-the-art descriptions of several currently available tests assessing the acquisition of signed languages and to outline the ongoing need for and challenges to the test development process.

<1> Educational and Research Contexts

The increasing diversity of the cultural and linguistic backgrounds of children in North American and European countries challenges traditional approaches to language testing (Johnston, 2007; Menyuk & Brisk, 2005). This diversity makes it difficult to determine the expected course of language development in bilingual children (Johnston, 2007). It therefore is important to develop parallel testing instruments for bilingual and multilingual children in order to measure their development in all their languages. 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, as increasing heterogeneity within Deaf communities has been reported in many countries (Christensen & Delgado, 1993; Gerner de Garcia, 2000).

Education for deaf children has always emphasized measurable outcomes of language and literacy learning, and the focus on bilingual and bimodal education for deaf children has highlighted the need for more information about early signed language development (DeLana, Gentry, & Andrews, 2007). Although assessment is a pivotal

component in educational programs serving hearing children, teachers in deaf education have typically relied on informal assessments in the form of naturalistic observations and anecdotal progress monitoring due to the limited tests supporting signed language assessment (Haug, 2005; Herman, 1998; McQuarrie, Abbott & Spady, 2012; Singleton & Supalla, 2011). In the current era of standards, accountability, and achievement testing, there is ever-increasing pressure to document learning outcomes (McQuarrie, Abbott, & Spady, 2012).

Accurate assessment can serve a variety of purposes, including determining the level of signed language proficiency when children begin school and the need to monitor progress. Children struggling to acquire language skills are often identified by professionals through assessment; therefore, identification of acquisition difficulties and strengths is yet another purpose of assessment. Assessment is also required for reporting purposes to inform parents and administrators of individual or group levels of functioning and rates of progress. These various purposes clearly identify the need for effective signed language assessment tools, and yet such tools are just beginning to be developed. The gap in the area of reliable and valid assessment measures of signed language acquisition, in comparison to the multitude of assessments available for spoken and written languages, leaves professionals working with deaf children without the necessary tools to assess, document, and track children's developing signed language competence. Therefore, another critical purpose for developing accurate signed language assessment measures is to emphasize the value the role that these skills play in deaf children's learning and facilitate teaching approaches that build on their strengths and abilities with visual language.

<2>Common Challenges

The need for assessments notwithstanding, there are legitimate challenges to the development of signed language assessment tools that contribute to the small number of tests currently available in this area. These challenges include the limited or lack of information regarding the acquisition of signed languages, the appropriateness of test formats based on spoken language tests, and the process for determining normative samples given the diversity of deaf children.

<3>Signed Language Acquisition Data

Language researchers have defined some key developmental milestones and acquisition patterns in the signed language development of young deaf children (French, 1999; Lillo-Martin, 1999; Newport & Meier, 1985; Schick, 2003). More information is available for some signed languages (for example, American Sign Language – ASL, and British Sign Language - BSL), and knowledge of the linguistic features of these languages and their relative grammatical complexity have been used to develop guidelines regarding the sequence of acquisition (Neidle, Kegl, MacLaughlin, Bahan, & Lee, 2001; Valli & Lucas, 1992).

In addition to the variability in knowledge of language acquisition among signed languages, there is also variability regarding the knowledge of specific language components (phonology, morphology, semantics, syntax, and pragmatics) of signed languages. Research into these components is primarily motivated by how specific signed language skills can contribute to the development of reading, or literacy in written language. Increasingly, researchers have explored the use of signed languages to promote the acquisition of written languages (see Chamberlain, & Mayberry, 2000). While several

correlational studies found positive relationships between ASL proficiency and reading proficiency (Hoffmeister, 1994, 2000; McQuarrie & Abbott, 2010; Padden & Ramsey, 2000; Prinz & Strong, 1998; Strong & Prinz, 2000) the exact nature of this relationship is not yet understood. To further address this question, developmental data regarding signed language phonology, vocabulary, and syntax have been used to create assessment measures that provide reliable indicators of these skills in young bilingual deaf children. The connection between signed language acquisition research and the development of practical assessment tools continues to be strengthened and extended across signed languages in the creation of important experimental and formal measures.

<3>Test Formats

In many ways, signed languages function similar to spoken languages, therefore, numerous assessments tasks used with spoken language can also be applied to signed language. A key difference, however, is modality and the need for visual vs. auditory stimuli. The increased availability of video formats and the accessibility of technology to play video has reduced the challenge of incorporating visual stimuli, and specifically signed language stimuli, within assessment measures.

There are several issues specific to signed languages that need to be taken into consideration when developing test items. The first issue concerns lexical item selection. Published high-frequency word lists or vocabulary lists for signed languages are not readily available, so simply determining appropriate lexical items suitable for young deaf children can be a challenge. Additionally, although many spoken language tasks involve visuals – pictures, wordless books, even videos or cartoons – some aspects of these stimuli can be overly distracting and are not designed for visual language users. In order

to overcome the challenges of test format and design, the process of signed language test development should always involve consultation with a panel of experts. These panels would consist of native signed language users who are researchers or specialist teachers in the particular signed language being assessed. The expert review panel, usually over a series of sessions, would conduct a content review of the test items to verify that the target sign stimuli, videos and/or pictures represent the constructs intended to be measured, and that any distracter items represent the most appropriate potential errors. Consultation with experts is a key component in the development of valid tests of signed language acquisition.

<3>Normative Samples

Identifying developmental problems in the acquisition of minority languages, whether signed or spoken, is challenging because norms for these populations often do not exist (Johnston, 2004). There is a lack of controlled elicited data from representative samples of native users of various natural signed languages upon which norms for competency can be established (Schembri, Wigglesworth, Johnston, Leigh, Adam, & Barker, 2002). The number of studies of signing deaf children's language development is limited and in the studies that do exist, the number of subjects is small. This is because only a minority of deaf children (less than 10%, Mitchell and Karchmer, 2004) can be considered native signers, with a normal experience of language acquisition from exposure to deaf parents who sign. For this reason, the general procedure for establishing signed language assessment involves initial pilot testing of items with native signers (children of deaf parents), followed by a broader normative testing process to include children with early exposure (before the age of 3 years) to the natural signed language.

To date, there is very little information on age-related knowledge of signed languages and its impact on learning. Studies of the relationship between specific areas of children's signed language abilities and areas of their spoken/written language literacy are important for determining the factors that predict reading ability, a crucial question for both educators of deaf students and researchers interested in bilingual acquisition issues. Tests of natural signed languages have tremendous practical value. The intent in developing formal assessments of signed language is to assist educators and researchers alike who work with deaf children in identifying children whose language is developing at age appropriate levels as well as those who are potentially at risk for language delay, language learning difficulties, learning disabilities, and classroom problems.

<1>Test Descriptions

Despite the common challenges of acquisition data, test format and normative sampling, several effective assessment tools of various signed languages have been established. We present descriptions of several such tests for ASL, BSL, DGS, and DSGS, including those that have been standardized as well as those still in development. This is not an exhaustive list of all available signed language assessments in these four languages, but rather a sample of tests that were presented as part of a symposium at the International Congress for Education of the Deaf (ICED 2015) that we, as authors, have developed in collaboration with each other and our research teams. In addition, the test descriptions have been summarized in Table 1.

<table 1 here>

<2>The American Sign Language Assessment Instrument (ASLAI) (Hoffmeister, Caldwell-Harris, Henner, Benedict, Fish, Rosenburg, Conlin-Luippold, & Novogrodsky 2014).

<3> History and purpose

The ASLAI is reliable and modeled on tests for spoken language development and tests of reading achievement, measuring conversational abilities, academic language knowledge, language comprehension, analogical reasoning, and metalinguistic skills. More specifically, it provides a measure of the relationship between specific areas of children's receptive signed language abilities and comparable areas of their English literacy skills. The results of the ASLAI are critical for determining factors that predict English reading ability. In addition to obtaining age related norms on the above language components, the ASLAI tasks can aid in identifying specific learning and/or language problems, which can lead to improved instruction and classroom settings. The ASLAI is designed to test deaf and hard-of-hearing students between the ages of 4 and 18 years. There are twelve subtasks in the total battery.

<3>Psychometric information

Reliability information for the ASLAI tasks comprising the vocabulary, reasoning, comprehension, and syntax domains indicates high reliability for all tasks.

Face validity for the ASLAI was established in two ways. First, the ASLAI tasks were developed by a team of native signers (deaf and hearing) whose first language was ASL. This ensured that the tasks used appropriate ASL forms and were suitable for the targeted constructs of the assessment. Second, all questions developed for the tasks were

piloted with native, deaf signers. Questions that did not have over 85% agreement were not used in the final version of the ASLAI.

Predictive validity was established for the vocabulary tasks in the ASLAI by determining how much variability on three different English reading and vocabulary assessments (the Stanford Achievement Test Reading Comprehension (SAT-RC) and Reading Vocabulary (SAT-RV) tasks, and the Measures of Academic Progress Reading subtask (MAP-Reading)) could be predicted by each of the ASLAI tasks. All ASLAI tasks predicted a significant amount of variability for the SAT-RC, SAT-RV, and MAP-Reading tasks. The overall results of the predictive validity of the ASLAI strengthen the use of the ASLAI as an appropriate measure of ASL age related skills, but also strongly demonstrates the relationship between ASL vocabulary knowledge and English (Hoffmeister, Henner, & Caldwell-Harris, 2015).

<3> Format/Platform

The ASLAI is a web-based assessment application. This allows for multiple participants to simultaneously take the assessment without the need for one-on-one administration. The web-based approach minimizes the resources necessary to assess large numbers of participants, and allows rapid testing and timely results.

The multiple-choice structure of the ASLAI has an advantage over criterion-based screening tools. Norm-based, multiple-choice assessments like the ASLAI remove the subjective component of scoring. Responses are either right or wrong, which increases the efficiency of data collection, analysis, and reporting which is helpful for schools and researchers.

The ASLAI testing platform is a proprietary design, developed from the ground up to ensure the security and confidentiality of all data collected from participants. Three components comprise the fundamental design of the ASLAI task platform: a) a stimulus window, b) up to four response windows, and c) a response review screen. The windows containing videos or images are shown sequentially, starting with the stimulus and followed by each of the responses in turn.

The testing procedure consists of five phases: a) the log-in, b) the instruction phase, c) the practice phase, d) the task phase, and e) the review phase. Students log-in to the ASLAI task battery using an individualized username and password combination which is maintained in case the participant is tested again in future years. Confidentiality and anonymity are maintained throughout the testing and analysis processes. During the instruction phase of the ASLAI, two types of instructions are presented: a) a general introduction to the ASLAI and how to interact with the testing platform, and b) instruction specific to each subtask.

During the third phase of the testing procedure, the practice phase, the instructions given in the instruction phase are reinforced and participants are given the opportunity to take up to five different practice questions reflecting the kind of questions that will be used in the task. The design of the practice interface (Figure 1) mirrors what is used in the task interface. For example, the practice interface for Vocabulary: Simple will have a picture stimulus and four responses. During the practice phase, participants receive feedback regarding whether or not they've chosen the correct response (correct responses are highlighted in green and incorrect ones in red).

<figure 1 here>

In the task phase of the ASLAI testing process, the actual task is executed. During this phase, participants are not given feedback on whether or not a response chosen is correct; rather, the task automatically advances to the next question after the participants select a response. Tasks in the ASLAI are in one of six formats: 1) picture to sign, 2) sign to sign, 3) picture to picture, 4) drag-and-drop sorting, 5) response-only (grammaticality judgment), and 6) video event to sign. Data collected during the task phase are securely sent to a central database, where performance is automatically scored and compared to other scores in the norming pool.

The review phase (Figure 2) is the final phase where participants are given an opportunity to review the responses they selected and change any answers if they so desire. During this phase, participants view a review screen that shows the stimuli from all task questions as well as the responses they selected. Participants are able to select questions and revise their answers before the final submission of task data.

<figure 2 here>

<3> Content/Design

The tasks in the ASLAI battery can be divided into four categories: 1) tests of vocabulary, 2) tests of reasoning skills, 3) tests of syntax, and 4) tests of ASL text comprehension. This section provides more detail about the tasks that compose the ASLAI test battery.

<4>Tests of Vocabulary

The vocabulary tasks in the ASLAI examine breadth and depth of vocabulary via antonymy, synonymy, and sign/word knowledge. These tasks require some level of metalinguistic judgment. Antonyms and Synonyms require students to make use of

metalinguistic knowledge in order to understand and identify differences and similarities among vocabulary items. Vocabulary in Sentences task is a higher-level vocabulary task, as participants must know both the meaning of the vocabulary item as well as its appropriate usage in different syntactic environments (sentences). In addition, the vocabulary items used as stimuli are taken from a pool of what we refer to as rare ASL vocabulary, that is, in-group vocabulary that is not often encountered unless one is intimate with the Deaf community. This task was developed for students aged 7.6 to 18 years old.

<figure 3 here>

<4>Test of Reasoning Skills: Analogies (24 questions)

The Analogies task is a test of classical language analogies in the sign to sign format. As seen below in Figure 4, the analogical stimulus sentence (*A is to B, as C is to what?*) is signed. Participants then view four possible signed lexical item responses, from which they are to select the correct one. Analogies contains a total 24 questions divided among six types of relationship categories: A) Causal, b) Purpose, c) Antonym, d) Noun-Verb Pairs, e) Whole to Part/Part to Whole, and f) Phonology.

<figure 4 here>

<4>Tests of Syntax

The third category of ASLAI tasks contains 27 items that examine knowledge of diverse aspects of ASL syntax. Nine different sentence types are represented: a) plain, b) conditionals, c) topic-comment, d) complement, e) relative clause, f) negation, g) rhetorical question, h) wh-question, and i) subject-object agreement.

The Classifier Category Sorting task uses a drag and drop format and is designed to measure knowledge of the ASL classifiers system. Classifier types represented in this task include semantic, handling, and size and shape specifiers.

The Real Objects and Plurals task measures knowledge of Verbs of Motion (VoM), Verbs of Location (VoL), classifiers, and pluralization processes in ASL. Scores indicate age-related knowledge of which classifier appropriately represents which object(s) (singular or plural), and how these classifiers forms function in VoM and VoL.

<4>Test of ASL Comprehension: ASL Text Comprehension (10 Questions)

The fourth and final category of tasks in the ASLAI contains one subtask: ASL Text Comprehension. It presents ASL texts (1 – 1.5 minutes in duration) and then asks five multiple-choice comprehension questions in ASL about that text. This task examines the ability of participants to extract both literal and inferential meaning from ASL texts. Participants view two ASL texts, responding to a multiple-choice selection of ten questions. This task was designed for students in all age groupings, with age-appropriate texts selected for each group.

<3> Reporting/Data

The ASLAI obtains age-related data from three different norming groups representative of the general population of deaf children: a) all deaf children, b) native signers, or Deaf Children of Deaf Parents (DCDP), and c) non-native signers, or Deaf Children of Hearing Parents (DCHP). For each ASLAI task norms were created using the means and standard deviation as they relate to age. Half standard deviations were used to compare ASLAI participants resulting in a more granular analysis of the language abilities relative to their peers. This is important especially at ages where participant

language skills are rapidly developing and any small change can cause gaps in language acquisition.

The strength and promise of the ASLAI has led schools to use the ASLAI as part of their yearly standardized testing battery. The ASLAI has been used with over 1500 students and is designed to support schools and programs serving Deaf children to identify those students who are performing as expected from those who are not acquiring or delayed in achieving ASL proficiency. Preliminary results based on this sample demonstrate that fluency in ASL, including knowledge of breadth and depth of ASL vocabulary, predicts English reading ability (Hoffmeister & Caldwell-Harris, 2014; Novogrodsky, Caldwell-Harris, Fish, & Hoffmeister, 2014; Novogrodsky, Fish, & Hoffmeister, 2014). Furthermore, the types of language errors fluent L1 users of ASL make are similar to the types of language errors fluent L1 users of English make (Novogrodsky, Fish, et al., 2014). Finally, the strength of the relationship between ASL abilities and English that we have found indicates that a positive relationship exists between ASL and English. Due to the large sampling population we have obtained for the ASLAI, we are confident these preliminary results are generalizable.

Achieving ASL proficiency as an L1 is critically important to learning English via print as an L2 (see Easterbrooks, Cannon, & Trussell, this volume; Mayer, this volume). The ASLAI provides us with a measure that is able to predict those students who may be having difficulty in learning English (L2). The ASLAI may also serve to separate those students who are having language learning problems displayed in both their L1 and their L2. This information has not been available for schools until now.

<2>BSL Receptive Skills Test (Herman, Holmes & Woll, 1999)

<3>History, purpose and target population

The British Sign Language (BSL) Receptive Skills Test (RST) was the first standardized measure available for any signed language. The test targets comprehension of selected aspects of BSL morphosyntax and was designed for child signed language users between the ages of 3-12 years. Under certain circumstances, the test can also be used with older children whose sign language skills are delayed or impaired or who have cognitive delays, although in such cases, the standard scores do not apply.

Use of the test enables professionals working with deaf children to make baseline assessments, identify language difficulties, and evaluate the outcomes of language therapy programs (Herman, Holmes, & Woll, 1998; Herman, 1998). The test provides an overall level of functioning which can be determined as age-appropriate or not (above/below, or significantly above/below average) and also a profile of errors to guide instruction about which grammatical structures students are struggling to understand. Children's results on this test can be compared with results for expressive language, using the BSL Production Test (Herman, Grove, Holmes, Morgan, Sutherland, & Woll, 2004). Both of these tests were key measures in a recent UK research programme seeking to identify and characterize language impairment in deaf signing children (Mason et al., 2010).

The BSL RST test is used widely in schools throughout the UK and has been used in its original or adapted form in a number of research studies with deaf children (e.g., Dammeyer, 2010; Davidson, Lillo-Martin & Chen-Pichler, 2014; Falkman, Roos & Hjelmquist, 2007; Jackson, 2001; MacSweeney et al., 2002; Mason et al., 2010; Sieratzki, Calvert, Brammer, David & Woll, 2001; Surian, Tedoldi & Siegal, 2010)

Tomasuolo, Valeri, Di Renzo, Pasqualetti & Volterra, 2012; Woolfe, Want & Siegal, 2002).

The test has been adapted into many different signed languages including German Sign Language (DGS, Haug 2011), American Sign Language (ASL, Enns & Herman 2011), Spanish Sign Language (LSE, Valmaseda, Pérez, Herman, Ramírez & Montero 2013), Italian Sign Language (Meristo et al., 2007) and Polish Sign Language (PJM, Kotowicz, personal communication 27th April 2013) among others. More recently, the DVD format has been re-developed into a web-based format, offering the possibility of updating test norms through use (Haug, Herman & Woll, 2014).

<3>Content/Format

The BSL RST has two components: the vocabulary check and the video-based RST. The original RST was presented on VHS and later updated to a DVD version. A web-based version is now available (www.signlanguagetest.com). The main test is presented on video, therefore minimal BSL skills are required by testers, although some BSL skills are needed to administer the vocabulary check.

The vocabulary check is optional and is designed to ensure that children understand the vocabulary used in the RST. It is particularly recommended for very young children, those who have had late exposure to BSL or children where language difficulties are suspected. Children complete the vocabulary check live using a simple picture-naming task that identifies signs in their lexicon that vary from those used in the RST. This is particularly important for languages such as BSL where there is much regional variation of signs. The vocabulary check takes approximately 5 minutes to administer.

The RST consists of 40 items, organized in order of difficulty. The items in the test assess children's comprehension of BSL morphosyntax in the following areas: (1) negation, (2) number and distribution, (3) verb morphology, (4) noun-verb distinction, (5) size and shape specifiers, and (6) handling classifiers.

The test procedure is explained by a deaf adult on the test video using a child-friendly BSL register. The test includes three practice items to familiarize the child with the test format. Children respond by selecting the most appropriate picture from a choice of four, previously presented in the accompanying color picture booklet and more currently, displayed on the computer screen. Repetition of test items is permitted for the practice items but not for the main test. An exception is for the very youngest age group (3-4 years), for whom a single repetition is allowed.

The test takes up to 20 minutes, depending on the age and ability of the child. Testing is discontinued after four consecutive failed items. Scoring is on a pass/fail basis. It is also possible to analyze a child's performance according to the grammatical features tested to identify strengths and weaknesses and targets for intervention.

<3>Psychometric information

The BSL RST was initially developed and piloted on 41 children (28 deaf and 13 hearing) aged 3;0 -11;6 years, all from native signing backgrounds. The revised and shortened test was subsequently administered to 135 deaf children within the age range 3-13 years located throughout the UK to establish test norms. For the standardization phase, children were included from both deaf and hearing families to represent the broader population of children who use BSL. Children previously diagnosed as having additional disabilities were excluded from the standardization sample. Additionally, children

scoring below one standard deviation on two subtests of the Snijders-Oomen Test of Non-Verbal Abilities (Snijders, Tellegen & Laros, 1989) were excluded from the sample. All tests were administered by a deaf researcher with fluent BSL and a hearing researcher with good BSL skills.

Test-retest reliability, split-half reliability and interscorer reliability were all investigated and reported for the revised task, showing the test to be psychometrically robust. Data were subsequently collected from an unselected sample of deaf children who used BSL throughout the UK for comparison with the standardization sample, reported in Herman & Roy (2006). The wider sample of children included many with additional learning difficulties and some older than the target test age, although these were excluded for the purpose of comparison with the original sample. Children in the new data set were found to achieve lower levels than those of the standardization sample, highlighting the importance of language monitoring and the need to improve language support provided in schools for deaf signing children.

<2>ASL Receptive Skills Test (Enns, Zimmer, Boudreault, Rabu, & Broszeit, 2013)

<3>History

The ASL Receptive Skills Test was adapted from the BSL Receptive Skills Test through a series of phases, including consultation with experts, development of new test items, videotaping of ASL stimuli, and re-drawing of picture responses. Two rounds of pilot testing were administered with native signing children (deaf children of deaf parents) to establish appropriate stimuli and distracter items and the accurate developmental ordering of test items. Following the first round of pilot testing (with 47 children in Canada and the United States), revisions were needed for 23 of the original 41

pilot test items, including changes to distracter drawings (11 items), signed stimulus sentences (4 items), and changes to both drawings and signed sentences (8 items). Four new items were added to assess understanding of the more complex structures of role shift and conditional clauses. Test items were re-ordered to more accurately reflect the developmental level of difficulty according to the number of children who passed each item.

The second round of pilot testing revealed that modifications to previous test items and the new test items made the test more challenging and more clearly distinguished children's skills at different ages. Analysis comparing age and raw score showed a significant correlation and high *r* value (r (34) = 0.821, $p < .001$). Final modifications included deleting three test items (considered redundant) and re-ordering test items to more appropriately reflect the developmental sequence of ASL acquisition (for more detailed information regarding the adaptation process, please see Enns & Herman, 2011).

<3>Purpose and target population

The purpose and target population are similar to the BSL RST but for ASL, that is, a measure of children's (ages 3 to 13 years) comprehension of ASL morphology and syntax.

<3>Content/Format

The content and format of the ASL Receptive Skills Test is similar to the BSL RST in that a vocabulary check (20 items) precedes the main test of 42 items. In addition to the 6 grammatical structures (number/distribution; noun-verb distinction; negation; spatial verbs; handling classifiers; and size-and-shape classifiers) assessed in the BSL

RST, the ASL RST also assesses the complex syntactic ASL structures of conditional clauses and role-shift. The original test format was revised by digitizing the picture responses and incorporating them into the test DVD (Figure 6). This eliminates the need for the picture book, and the child is not required to shift eye gaze between the computer screen and the picture book, thus reducing distractibility errors. A web-based version is currently in development (www.signlanguagetest.com).

<figure 5 here>

<3>Psychometric information

The ASL RST was administered to 203 children throughout Canada and the United States for standardization. Deaf children from hearing families were included in the standardization sample; however, only if they had been exposed to ASL by age 3 or younger. All 203 children were deaf and had a non-verbal IQ of 70 or above (or where formal testing was not available, were determined to be functioning within the average range intellectually by school personnel). There were 77 native signers and 126 non-native signers (acquired <3 years old), 106 females and 97 males, and the ages ranged from 3 to 13 years. Testing took place in the children's schools and was administered by deaf and hearing researchers with fluent ASL skills. We recognize that our sample of 203 children is limited in how accurately it represents the overall population of deaf children, and for this reason future research will involve additional testing and data collection to expand our sample. However, several statistical analyses of the standardization data did reveal that the test was reliable (showed internal consistency) and was a valid measure of developmental changes in ASL skills.

<2>German Sign Language (Deutsche Gebaerdensprache: DGS) Receptive Skills Test (Haug, 2011)

<3>History

The DGS Receptive Skills Test was adapted in a series of steps from the original BSL RST to German Sign Language (see also Table 2). The first revisions included changing some of the images of the BSL test to adapt it to the German context, for example, the British red and round mailbox was replaced by a yellow German mailbox. After reviewing the literature on DGS research, the items were adapted into DGS. Most of the linguistic structures that occur in the BSL test could also be represented in DGS, while others are not part of DGS morphology (i.e., noun-verb derivation). Ten additional items were created as potentially not all items would work equally well in DGS as in BSL. The original adapted items followed the same order as the BSL items, followed by the ten additional items. After the pilot study, the test was administered to 54 deaf children aged 3;9 to 10;10 years.

<table 2 here>

<3>Purpose and target population

The purpose and target population are identical to the BSL and ASL tests, i.e., targeting deaf children (ages 3-11 years) to evaluate their comprehension of DGS morphology and syntax.

<3>Content/Format

The current version of the DGS RST consists of a vocabulary check (22 items) and 49 test items, representing DGS structures of number and distribution, negation, spatial verbs, handling classifiers, and size-and-shape classifiers. The first version of the

DGS RST was delivered on a laptop, using a specially designed stand-alone application. A web-based version is currently in development (www.signlanguagetest.com).

<3>Psychometric information

The DGS RST has been tested on 54 deaf children in Germany; 34 of these children had at least one deaf parent, the remaining children had hearing parents (e.g., Haug, 2011, 2012). Based on an item analysis, 10 of 49 items needed to be removed or revised, and many items were "too easy", therefore, more difficult items are currently being developed.

Evidence for reliability was established across all 49 items through statistical analysis, and a significant correlation between chronological age and raw score determined that the test was sensitive to age differences. These results applied to the sample as a whole as well as for both sub-samples separately: deaf children of deaf parents, and deaf children of hearing parents. Currently the authors are looking into funding for a planned norming study in 2015/16. Normative data will enable educators to determine if children are acquiring DGS age-appropriately and/or to establish goals and effective teaching strategies for those who are demonstrating delays or disorders in their language development.

<2>BSL Production Test (Herman, Grove, Holmes, Morgan, Sutherland & Woll, 2004)

<3>History, purpose and target population

The BSL Production Test (Narrative Skills) targets narrative skills and use of BSL grammar based on a narrative recall task. The test is designed for child signed language users between the ages of 4-11 years, since narrative skills develop during these years. Using a narrative sample is ecologically valid, since the language produced is more

naturalistic than that found in other types of assessment (e.g. sentence repetition). Furthermore, narrative skills are sensitive to language impairments (Norbury & Bishop, 2003) and correlate with literacy development (e.g. Reese, Suggate, Long & Schaughency, 2010) in hearing children.

To use the BSL PT, testers must have advanced fluency in BSL and complete a training course to learn the coding system. The test is used in schools throughout the UK and has been used in research studies with deaf children to effectively serve as a baseline measure of BSL abilities, to assist in the assessment of language and learning difficulties, and to monitor progress following intervention (e.g. Herman, Rowley, Mason & Morgan, 2014; Kennedy et al., 2006; Mason et al., 2010). To date, the test has been adapted into Australian Sign Language (Auslan: Hodge, Schembri & Rogers, 2014) and plans are underway to develop an ASL (see following section) and LSE version (Perez Martin, personal communication, 29th October 2014). In addition, a spoken English version of the test (Jones et al., 2015) is currently under development that will enable practitioners and researchers to use the same test to assess a deaf child's spoken and signed narratives.

<3>Content/Format

The BSL PT is a narrative recall task based on children watching a two-minute language-free video presented on TV/computer. The video features a boy and a girl acting out a series of events without communicating to each other in either signed or spoken language. Children are told that they will watch a video and then tell the story to a deaf BSL user who has not seen the video. If the tester is not a fluent BSL user, it is recommended that a fluent BSL user be involved when children tell the story to ensure narratives are delivered in BSL rather than English-based signing. The child may watch

the video a second time if they wish and the child's story is video recorded for later analysis.

The child then answers questions targeting story comprehension and inferencing skills. The questions are prerecorded for presentation on video; however, for some children (e.g., very young children or those with attention or language processing difficulties), testers may need to present questions live. Responses to questions are video recorded for later analysis.

Scoring is based on spontaneous recall of the story without prompts and children's responses to questions. Samples are coded for three aspects:

(i) Narrative content. Children's narratives are coded for the explicit mention of 16 narrative episodes (maximum = 16). The score for responses to questions (maximum = 6) is included within the narrative content score.

(ii) Narrative structure. Based on a high point analysis (Labov & Waltesky, 1967), narratives are coded for orientation; complicating actions; climax; resolution; evaluation and sequence (maximum = 12).

(iii) BSL grammar. Correct use of morphological inflections is coded for spatial verbs including classifiers, agreement verbs, manner inflections, and aspectual inflections (maximum = 30). Narratives are also rated for mastery of role shift (rated 0-4).

The test takes up to 10 minutes to administer, depending on the age and ability of the child. Following analysis of a child's story and responses to questions, the raw scores obtained can be converted to percentiles. It is also possible to analyze a child's performance according to the narrative and grammatical features tested to identify strengths and weaknesses and identify targets for intervention. The test manual provides

details of the aspects of BSL grammar included in the test and information about the development of narratives in deaf and hearing children, including those with special educational needs.

<3>Psychometric information

The BSL PT was initially developed and piloted on the same 41 native signers used to develop the BSL RST. Following piloting, the test was administered to 75 deaf children (34 boys) from deaf and hearing families within the age range 4-11 years located throughout the UK to establish test norms. As for the BSL RST, children with additional disabilities or who scored outside the normal range on non-verbal measures were excluded. All tests were administered by a deaf researcher with fluent BSL skills.

Test-retest reliability, split-half reliability and interscorer reliability were all investigated and found to be good. Concurrent validity, explored by comparing scores on the BSL RST and BSL PT, was found to be high.

<2>ASL Production Test (Enns, Boudreault, Zimmer, Broszeit & Goertzen, 2014)

<3>History, purpose, and target population

The ASL Production Test is an adaptation of the BSL Production Test. Since the BSL Production Test involves a narrative elicitation task through the use of a language-free story on video (Spider Story), it has good potential for use in any language. Essentially, the analysis and scoring of narrative content (events in the story) and narrative structure (story development) are the same across languages. So the adaptation into ASL was specifically focused on the grammar analysis. The BSL grammatical categories of spatial verbs, agreement verbs, aspect, manner, and role shift fit well with ASL grammatical

categories therefore adapting the scoring to the specific features of how these are marked in ASL was quite straightforward.

The next phase in the project was to create additional versions of the test, or parallel video-based stories that would elicit comparable narratives. Having alternate elicitation videos allows for re-testing students without them becoming familiar with the story over repeated viewings. It also allowed for updating the original Spider Story, from the BSL test, to incorporate American cultural features and improve the video quality. Throughout this process, however, it was essential to keep the narrative content, structure, and grammar similar across all three versions of the test videos. We used the basic narrative structure of the Spider Story and aligned the two new stories, "Home Alone" and "Tiffany's Breakfast" to this framework. Each story consists of similar events to the Spider Story (a series of back-and-forth interactions between protagonist and antagonist) but with slightly different settings, characters, and consequences. There are also parallels between the objects and actions in each of the stories that allowed for opportunities to elicit the same kinds of grammatical structures (spatial verbs, agreement verbs, aspect, manner, and role shift). In addition, the stories needed to be enjoyable and engaging for children so they would remember them and be interested in retelling them.

The third phase of the adaptation process, pilot testing of the adapted and new test versions with a sample of typically developing native ASL signers aged 4 to 12 years, is currently underway. This information will provide valuable feedback regarding the effectiveness and reliability of the scoring guidelines, as well as the equivalency across the three test versions. Once the necessary revisions are made based on the results of pilot testing, the final phase of standardization on a larger normative sample will be conducted.

<3>Content/Format

The test content and format is similar to the BSL Production Test (Herman, et. al., 2004), in that the child watches the video elicitation and then spontaneously re-tells the story and answers three comprehension questions. The child's responses are video recorded and analyzed according to specific scoring guidelines. The goal for the ASL Production Test is to create both an online training process (mandatory for all testers), and an online scoring system (to simultaneously view the scoring rubric and the child's narrative). The required training process is needed to ensure that all testers have the skills to score the narratives accurately and consistently. As previously described, the ASL Production Test will also have three possible test versions (language-free video stories) to elicit comparable narratives from children over time.

<3>Psychometric Information

The pilot testing phase is currently being conducted, therefore testing results are not yet available. The reliability of the ASL Production Test will be investigated using intra- and inter-scorer comparisons, test-retest analyses and a measure of internal consistency. Test validity will also be investigated by comparing scores from the same children on both the ASL Receptive Test and the ASL Production Test, as well as other measures of academic skills and performance (non-verbal IQ, reading comprehension). The results of the analyses procedures will determine the final version of the test that will be used for larger scale standardization (planned for 2016). Standard scores will be determined through statistical analysis of the collected norms.

<2>BSL Non-Sign-Repetition Test (NSRT) (Mann, Marshall, Mason, & Morgan, 2010)

<3>History

Research into the development of signed language phonology is often supported by only partial linguistic descriptions of signed languages and far fewer studies of the acquisition of those languages. Because most signs are only one syllable long (Brentari, 1998), it is not possible to manipulate the length of a sign. Therefore, adapting a non-word repetition paradigm for signed language offers the possibility of manipulating signs with regards to their phonological complexity along two parameters – handshape and movement. This methodology makes it possible to investigate the perception, retention and production of novel phonological forms, in both deaf and hearing children.

<3>Purpose

The Nonsense Sign Repetition Task (NSRT) for British Sign Language (Mann et al., 2010) assesses signing deaf children's ability to repeat nonsense signs of differing phonetic (i.e., handshape and movement) complexity in British Sign Language (BSL). The test is based on a pilot by Marshall, Denmark, and Morgan (2006).

<3>Target population

The NSRT was developed for signing deaf children ages 3-11 years. As part of the norming study, two groups of children were tested: (1) deaf children who acquired BSL as a first language (3-11 years old) and (2) hearing children with no prior experience/exposure to signing (6-11 years old). The first group consisted of 91 deaf children, who were divided into three age groups: 3–5 years old (N = 26); 6–8 years old (N = 26), and (3) 9–11 years old (N = 38). Fourteen of these children had deaf parents, and the remaining children had exposure to BSL from nursery school. The second group consisted of 46 hearing children, divided into two age groups: 6-8 years old (N = 23) and 9-11 years old (N = 23) (Mann et al., 2010).

<3>Content/Format

The chosen methodology is based on the non-word repetition methodology used in spoken language acquisition research (e.g., Dollaghan & Campbell, 1998; Gallon, Harris, & van der Lely, 2007). Items of the Nonsense Sign Repetition Test for BSL consist of items that are phonotactically possible, but do not carry any meaning in BSL. These items differ with regard to their phonetic complexity along two phonological parameters: handshape and movement. For each parameter, items are either phonetically "simple" or phonetically "complex". With respect to handshape, "simple" handshapes are the four unmarked BSL handshapes, labeled here as "B", "5", "G", and "A" (Sutton-Spence & Woll, 1999) and "complex" handshapes are all marked handshapes. With respect to movement, the test developers define just one movement, whether internal or path, as "simple", and two movements (i.e., internal and path combined) as "complex". The NSRT consists of a total of 40 items that are distributed equally across different levels of complexity. All items were modeled by a deaf signer and presented via a computer format.

Each child is tested individually and the test takes about 10-20 minutes. Before the actual test, instructions are presented on video by a deaf native signer and followed by 3 practice items. Each item is presented only once. The 40 items are shown in blocks of 10 items, with a short break between each block during which the child is shown a short cartoon (Figure 6).

<figure 6 here>

The responses are coded/scored according to whether the overall response was correct, whether any errors were made on the phonological parameters, and additionally whether one of the movements in a movement cluster was deleted.

<3>Psychometric information

The NSRT was developed and piloted on 91 deaf children with exposure to BSL from very early on through nursery school, including 14 children with deaf parents. Children with additional disabilities were excluded. All tests were administered by a hearing researcher with fluent BSL skills.

Parallel-forms reliability and inter-rater reliability were all investigated and found to be good. Content validity in the form of feedback given by three native signers resulted in removal of any signs that were not suitable. Concurrent validity was explored by comparing test takers' scores on the NSRT to a hearing control as well as to their performance on a fine motor skills task. All showed good values (see Mann et al., 2010, for more detailed information).

<2>American Sign Language Phonological Awareness Test (ASL-PAT) (McQuarrie & Spady, 2012)

<3>History

The ASL-PAT represents an adaptation of well-known spoken language psycholinguistic paradigms. It is designed to measure knowledge of the sublexical properties of sign formation (i.e., handshape [H], location [L], and movement [M]). McQuarrie's (2005) receptive-based phonological similarity judgment task (for ages 9 to adult) was used as a prototype in developing a downward extension of the measure suitable for 4-8 year old children (preschool through second grade) because performance

on the earlier measure reliably identified students' awareness of the sign language phonological segments that comprise signs; performance on this sign segmenting task was also correlated with success in English word-reading and reading comprehension (e.g., McQuarrie & Abbott, 2013; McQuarrie & Abbott, 2010; see review in McQuarrie & Parrila, 2014). In the first phase of test development an initial pool of test items was vetted by a team of native ASL users who were knowledgeable about child language development and were able to suggest representative content. The items were then pilot-tested on a group of twelve deaf children (four children in each of the youngest age categories). The purpose of the pilot test was to evaluate the feasibility, usefulness, and usability of the test items, and to examine the effectiveness of the instructions, items, and item delivery method. Only items that showed at least a 95% agreement among the young deaf respondents were retained in the item pool. Following the initial test item pilot and subsequent item revisions, the revised test items were again piloted with an additional group of young deaf children in order to examine response patterns, difficulty level of items, and how well test items discriminated among various groups of deaf children (i.e., early vs. late sign exposed children). A final round of pilot testing with a sample of typically developing native ASL signers aged 4 to 8 years is currently underway. The final version of the ASL-PAT will be optimized by including only the items that best predict phonological awareness in ASL and are most sensitive to developmental differences in phonological awareness (see McQuarrie, Abbott & Spady, 2012 for more detailed information regarding test development and design). The final phase of standardization on a larger normative sample will begin in Fall 2015.

<3>Purpose and target population

The ASL-PAT is being developed for use with signing deaf children ages 4-8 years (preschool through second grade). The aim is to develop a signed language phonological measure that is sensitive enough to discriminate young children's phonological knowledge based on age, and to distinguish native and late-learners of ASL.

<3>Content/Format/Test Platform

The ASL-PAT measures the ability to identify phonological similarity relations in signs under three comparison conditions:

1. signs with three shared parameters (H + M + L);
2. signs with 2 shared parameters (H + M; L + M; and H + L);
3. signs that share a single parameter (H, M or L).

The ASL-PAT is a web-based assessment application. Similar to the ASLAI, multiple individual users can access the assessment at the same time and all individual user responses are uploaded to a central database in real time. The testing procedure consists of five phases: a) log-in and background demographic questionnaire, b) vocabulary check, c) instruction video, d) practice trials, and e) test block. The testing takes about 10-15 minutes for each test-taker.

<4>Log in

An identification number is assigned to each user on log in. A brief questionnaire including background information (e.g., date of birth, gender, age of onset, age of exposure/acquisition, use of hearing technologies, age of implantation etc.) is completed online by the tester.

<4>Vocabulary check

The test begins with a vocabulary check in the form of a picture dictionary presented as a 5 X 5 grid picture display. Children are required to sign (name) each picture. If a child is uncertain or unable to generate a sign for a picture item, a video prompt of the sign is available by clicking on the picture. Prompted items are subsequently added to the end of the picture display and retested without the video prompt prior to beginning the test. It is essential that children know the vocabulary associated with the test pictures prior to taking the test.

<4>Instruction Video, Practice Trials, Test Block

Video instructions are presented in ASL by a deaf adult signer (Figure 7), followed by 7 practice trials and 28 test items. Feedback is provided on the practice items, no feedback is provided on test items. Each practice and test item consists of a signed cue (video) with three picture items below representing the target/phonological match and two distracter items. Test-takers are required to select the picture that matches the cue along the phonological parameter(s) tested (Figure 8).

<figure 7 here>

<figure 8 here>

<4>Scoring

The online database records accuracy (correct match -1; incorrect match - 0) and error response choice for each test item. Overall test performance scores are determined by the number of correct responses out of 28. Reaction/response time data is also recorded.

<3>Psychometric information

The ASL-PAT is still in development and psychometric information is not yet available. Preliminary findings from pilot studies offer support for (a) the plausibility of

assessing signed language phonological awareness by targeting key phonological parameters identified in the literature, and (b) the potential of web-based test delivery using a dynamic/video presentation format. In addition, McQuarrie & Enns (2015), using multiple single-case studies incorporating a multiple probe across skills design, found a clear functional relation between explicit instruction in ASL phonological awareness and increases in sign vocabulary and print vocabulary learning in young deaf dual language learners. These results were confirmed by changes on the ASL-PAT that was administered at the beginning and at the end of the study. Upon completion of the test development project and associated validation studies, the ASL-PAT will be available for teachers and clinicians to provide diagnostic information on children's sign phonological development and to identify children who lack explicit sign phonological knowledge. The test is expected to give reliable indicators of the development of ASL phonological awareness in young bilingual deaf children which may allow educators to establish targeted phonological learning objectives and plan effective sign phonological instructional interventions for bilingual deaf students.

<2>BSL Vocabulary Test (BSL-VT) (Mann & Marshall, 2012)

<3>History

Most standardized assessments of children's vocabulary draw on the mapping between phonological form and meaning. Typically the task involves presenting the phonological form (word) and requiring the test-taker to select a picture that matches its meaning from a set of three or four (PPVT, Dunn &Dunn, 1997), or providing a picture and the test-taker must produce the phonological form that matches the meaning (EOWPVT, Brownell, 2000). These assessments are limited in that they only measure one level of

vocabulary knowledge, for example meaning recognition, or form recall. These task limitations along with the lack of standardized vocabulary assessments for signed language motivated the development of a set of vocabulary tasks to assess different levels of deaf children's vocabulary knowledge in BSL. One advantage of having an assessment that provides more detailed information about a child's different levels of vocabulary knowledge is the impact it can have on guiding and improving intervention (Mann, Roy, & Marshall, 2013). In addition, because the BSL-VT enables identical items to be compared across more than one task, it possible to identify unusual language profiles as demonstrated in Mann, Roy, and Marshall (2013). In the Mann et al. study, the BSL-VT was administered to a larger, more diverse deaf sample, including deaf children with additional needs such as attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), dyslexia. While there was no significant effect of additional needs on vocabulary performance, an unusual response pattern was noted in one child with ASD which was consistent with reports of autism in spoken languages. This stresses the value of studies that assess vocabulary development in deaf children with a wide range of additional disabilities and might contribute important information about possible effects of disabilities on word learning commonly found within the group of deaf language users (Mann, Roy, & Marshall, 2013). In this context, the BSL-VT is particularly valuable due to its unique format.

<3>Purpose

The purpose of the web-based BSL-VT is to assess deaf children's vocabulary knowledge in British Sign Language (BSL) "by specifically measuring the degree of strength of the mappings between form and meaning for items in the core lexicon" (Mann

& Marshall, 2012, p. 1031). One particular aim is to investigate whether there is a hierarchy of difficulty for these tasks, and therefore whether BSL vocabulary acquisition proceeds incrementally as is the case for spoken languages.

<3>Target Population

The target population for this test is signing deaf children between the ages of 4 - 15 years. At this point, data have been collected from 67 children. Twenty-four deaf children from five programs that use BSL as the language of instruction participated in the pilot study (Mann & Marshall, 2012). Of these 24 children, 12 were male and their average age was 11;2. All of the participants had a hearing loss of >70 dB in their better ear. They were either native signers or strong signers, who all used BSL as their preferred language/means of communication.

In a follow-up study, an additional 43 children were assessed, resulting in a total of 67 deaf children (37 male, 30 female) aged 4-17 years (Mann, Roy, & Marshall, 2013). One difference from the pilot study was that the newly added participants had more variable BSL skills and also included children from diverse language learning backgrounds and children with additional needs. The goal of this study was to investigate if some key variables in deaf signing children such as parental hearing loss and additional needs affect deaf children's vocabulary knowledge in BSL (Mann et al., 2013). Average scores for three of the four tasks were reported.

<3>Content/Format

The web-based BSL-VT consists of four tasks to assess different degrees of vocabulary knowledge: (1) meaning recognition (test-taker sees a pre-recorded BSL sign followed by four pictures, and must select the picture that corresponds to the meaning of

the signed prompt); (2) form recognition (test-taker sees a picture, followed by four pre-recorded BSL signs and must select the sign that matches the meaning of the picture prompt); (3) meaning recall (test-taker sees a pre-recorded BSL sign and must generate another BSL sign with an associated meaning); and, (4) form recall (test-taker sees a picture and must produce the corresponding BSL sign). Each task consists of 120 items. The test draws on a model for second language learning (Laufer, Elder, Hill, & Congdon, 2004; Laufer & Goldstein, 2004) where the same items are used across all tasks. The test includes two receptive and two production tasks. The two receptive tasks (meaning recognition, form recognition) use a multiple-choice format and can be self-administered. The two production tasks (meaning recall, form recall) require an administrator who scores each response based on four options and also documents the given response in a text box on the computer screen, using English glosses.

The items for the test were selected from several sources: (1) a BSL norming study (Vinson, Cormier, Denmark, Schembri, & Vigliocco, 2008), (2) receptive vocabulary test for German Sign Language (Bizer & Karl, 2002a; Bizer & Karl, 2002b), (3) commonly used, standardized, English vocabulary tests and (4) feedback from deaf and hearing researchers and teachers who collaborated with the authors during the item development (Mann & Marshall, 2012).

The order of the items in each set is randomized every time someone takes the test. The items of the test belong to the grammatical categories of nouns, verbs, and adjectives. Based on research for spoken language acquisition which shows that the first acquired words are nouns, the ratio across the three grammatical categories is 8:1:1 for children < 10 years old and 6:2:2 for older children (> 10 years old). Signs known to have

regional variations (including colors and numbers) were excluded from the test. It was important to develop effective distracters for the two receptive skills tasks involving multiple-choice responses. These tasks include four types of responses that are presented randomly for each item: (1) the target, (2) a phonological distracter, (3) a semantic distracter, and (4) a visual or an unrelated distracter. Signs that were known to be iconic (e.g., body parts, animals, numbers) were excluded as much as possible, even though they are commonly used in spoken English vocabulary tests (Mann & Marshall, 2012).

<4>Procedure

The test is presented individually for each child by a signing tester, preferably a deaf native signer (Figure 9). Both receptive tasks can be self-administered by the test-taker, depending on age and familiarity with a computer and mouse. For the two production tasks, the test-taker produces responses in BSL and the tester enters the answer as English gloss during test administration. All results are saved automatically in the database on the Web server. The four tasks are completed in two sessions, and each session includes two tasks. There should be at least one week between the first and the second session to minimize learning. Each session takes about 30 minutes. Before the tasks start, the test-taker sees pre-recorded instructions in BSL, which can be elaborated on by the tester for younger children, and is given the chance to practice on two items.

<figure 9 here>

<4>Scoring

The responses for the two receptive skills tasks are scored as ‘1’ for correct and ‘0’ for incorrect. For the production tasks, four answer choices are provided for the tester to code answers live. For the form recall task, these scoring choices are (1) correct sign

(scored as '1'), (2) partially correct sign (scored as '0.5'), (3) wrong sign/different sign (scored as '0'), and (4) do not know (scored as '0'). For the meaning recall task, they are: (1) categorical response (scored as '1'), (2) non-categorical response (scored as '0.5'), (3) different/unrelated response (scored as '0'), and (4) do not know (scored as '0'). These scores are presented in codes (e.g., CS=correct sign) so as not to affect the test-taker's motivation. In addition, the English gloss is entered in a textbox below the coded answer as an added measure of reliability and in lieu of video recording the responses which would render the task less time efficient for practitioners.

<3>Psychometric information

The BSL-VT was developed and piloted on 24 deaf children, including deaf children with deaf parents and those recommended by teachers as strong signers. Following piloting, the test was administered to 43 deaf children from deaf and hearing families within the age range 4-17 years located throughout the UK, including children with additional needs. All tests were administered by a hearing researcher with fluent BSL skills.

Reliability was established in the form of inter-rater reliability and found to be good. Content validity was established based on feedback from a panel of deaf and hearing experts. Construct validity, explored by correlating test takers' performance on each of the four tasks and age, was found to be high. Concurrent validity was explored by comparing scores on the BSL-VT to Non-verbal IQ (i.e., Raven's Progressive Matrices) and controlling for age. Findings showed moderate correlations for all tasks (For more detailed information, see Mann, Roy, & Marshall, 2013).

<2>Web-based ASL Vocabulary Test (ASL-VT)(Mann, Roy, & Morgan, 2015)

<3>History

There are currently no standardized ASL vocabulary tests available that measure strength of vocabulary knowledge. The ASL-VT was adapted from the BSL-VT. It was developed as a baseline measure for a larger study investigating how deaf children respond to mediated learning in ASL, in addition to investigating the reliability and validity of the adapted measure.

<3>Purpose and target population

The purpose of the ASL-VT is to assess receptive and expressive vocabulary knowledge of signing deaf children in the USA (and other countries using ASL). The target population is deaf children aged 6 -10 years. At this point, data have been collected from 37 children, including deaf children with deaf parents, deaf children with hearing parents, and deaf children with additional needs. The findings, to date, indicate a similar hierarchy regarding level of difficulty for the four tasks as was found in the BSL study. Specifically, deaf children perform highest on the meaning recognition task, followed by the form recognition task and experience more difficulties with the two recall tasks – form recall and meaning recall. Our work on the ASL-VT is still in progress and awaits standardization on a larger sample.

<3>Content/format

The test uses the same format as the BSL-VT, including two receptive (multiple-choice) tasks and two production tasks (Figure 10). 66 items were translated directly from the BSL-VT and 14 revised or new items were added for a total of 80 items (this is fewer than the BSL-VT due to the smaller age range of participants in U.S. pilot).

<figure 10 here>

<3>Psychometric information

The ASL-VT was developed and piloted on 20 deaf children, aged 6-10 years all of whom came from deaf families. All tests were administered by a deaf native signer with fluent ASL skills.

Reliability was established in the forms of internal consistency, inter-rater agreement and item analysis. Content validity was established based on feedback from an expert panel, which included deaf native signers and rating scores from teachers. Construct validity was explored by correlating test takers' performance on each of the four tasks and age, analysis of differences between participants' performances across tasks, and comparison of deaf children's performance on the two receptive tasks to age-matched hearing children with no previous knowledge of sign. Concurrent validity was explored by comparison of ASL-VT scores to performance on the ASL-Receptive Skills Test (for more detailed information, see Mann, Roy, & Morgan, 2015).

Our similar findings for ASL and BSL suggest that the underlying construct, that is, vocabulary knowledge has different degrees of strength based on the mapping between form and meaning of signs, holds true for these two sign languages and indicates that the tasks would be applicable to other signed languages as well.

<2>Swiss German Sign Language (Deutschschweizerische Gebaerdensprache; DSGS) Sentence Repetition Test (Haug, Notter, Girard, & Audeoud, 2015)

<3>History

A survey among schools for the deaf in German Switzerland (Audeoud & Haug, 2008) confirmed the need for an educational assessment of signed language – a need that has also been confirmed in other international studies (e.g., Germany: Haug &

Hintermair, 2003; UK: Herman, 1998; USA: Mann & Prinz, 2006). In German Switzerland, where one of the three Swiss signed languages is used (Boyes Braem, Haug, & Shores, 2012), no standardized and normed test of DSGS exists. To meet this need, a research team at the University of Applied Sciences of Special Needs in Zurich (HfH) applied for funding to develop a sentence repetition test (SRT) for DSGS. The test was developed by a team of deaf and hearing researchers in cooperation with deaf signed language instructors experienced in working within a school context.

<3>Purpose and target population

The purpose of the DSGS Sentence Repetition Test (DSGS-SRT) is to develop a global measure of DSGS development that is easy to administer and score. The SRT-DSGS provides information on different linguistic levels, such as phonology, morphology, and syntax. Since only one study of DSGS acquisition (Fosshaug, 2010) is available, developing a test for DSGS development posed a methodological challenge. It was necessary to rely on available adult data to develop the reference measure.

The target population for the DSGS Sentence Repetition Test is signing deaf children between 6-12 years of age, with either deaf or hearing parents.

<3>Content/Format

Sentence repetition tests from other signed languages, such as American (Hauser et al., 2008), German (DGS; Kubus & Rathmann, 2012), British, children's version (Cormier et al., 2012), Swedish, and Italian Sign Language[i], served as templates for the DSGS-SRT development. Additionally, sentences from "e-kids", an online portal with DSGS teaching materials for children from the Swiss Deaf Federation[ii] were used as well as developing new sentences.

Since the ASL and the DGS version targets adult signers, the sentences that were adapted for DSGS were changed in regard to (1) avoiding lexical variations, (2) adapting the signing style to be more child-appropriate, and (3) matching children's experiences (e.g., "The last time I was on vacation was seven years ago" does not match the experience of a 6-year old child). Lengthy sentences that are more appropriate for adults were removed and specific DSGS linguistic features were added. The first version of the DSGS-SRT consisted of 76 sentences of increasing length and complexity. The panel of experts provided feedback on the initial version of the test and this resulted in removal and revision of some sentences. This process reduced the number of sentences to 60. The panel of experts met regularly during the life of the project to provide ongoing feedback.

The DSGS-SRT is presented on video, embedded in PowerPoint, and administered with a laptop computer. Following pre-recorded instructions in DSGS, three practice sentences are presented, followed by the test sentences (Figure 11). The test participants watch a sentence and then repeat it as accurately as possible. The test participants were video-recorded with the build-in webcam of the laptop on which the test was presented. The tester was sitting opposite and slightly to the right side of the test participants.

<figure 11 here)

After training two deaf testers, a pilot study was conducted with three deaf adults and three children (CODAs; ages: 7;5 – 11;11 years old). Besides feedback on the test procedure (test instructions), 20 additional sentences were removed when they were either mastered/not mastered at all by the pilot test participants. Also based on the feedback from the deaf adult participants and testers, some sentences were revised.

The 40 remaining sentences were video-recorded in a professional video studio. For the main study 50 deaf children 5-17 years old and 15 deaf adults were tested. The results are currently scored with a newly developed scoring tool. This scoring tool is similar to that developed for the BSL SRT for children (Marshall et al., 2014).

<3>Psychometric information

The investigation of the test's psychometric properties is currently underway.

<1>Discussion

The primary purpose of assessment is to inform educational decisions and instruction. With accurate tools to assess signed language competence, educators are able to establish baseline measurements and determine progress, identify students with language delays or disorders, and evaluate the outcomes of classroom or individual therapy programs. Researchers also have the tools to compare populations and provide consistency across studies. It is important that practitioners and researchers work together to develop signed language interventions that are based on valid and reliable test results in order to improve deaf children's language proficiency. The assessment tools described here clearly demonstrate that significant gains are being made towards establishing effective measures to enhance deaf children's acquisition of signed languages. Throughout the discussion of test development, two issues arose repeatedly and require further mention. These issues are adapting tests for use in other signed languages and implementing new technologies for test presentation.

<2>Test Adaptation

The process of adapting tests from one signed language to another requires careful consideration of the linguistic differences that exist between the two languages; however,

limited cross-linguistic research related to signed languages can make this a challenging task (Mason, 2005). These challenges are illustrated by Haug and Mann (2008) through examples involving differences in the categorization of linguistic features, lexical differences, and morpho-syntactic issues. Cultural issues also play a part in test adaptation. This can be as simple as pictures depicting the size, color, and shape of a British vs. German mailbox or as complex as a story involving the experience of obtaining a driver's license, which is common in America but not in Switzerland (Haug & Mann, 2008).

The decision of whether it is advantageous to adapt an existing instrument that has already been tested and standardized must be considered within the framework of evaluating the linguistic and cultural differences between the original and target languages. If test developers determine that these differences can be overcome by modifications to pictures, distracter items, or stimuli, then the adaptation process is considered worthwhile because important test development decisions have already been evaluated. For example, the BSL RST was based on what was known about signed language acquisition and highlights grammatical features identified in the research as important indicators of proficiency, such as verb morphology and use of space (Herman, Holmes, & Woll, 1998). Considering that many signed languages share these important grammatical features numerous test items are relevant in signed languages other than BSL, as evidenced by the ASL and DGS versions. Another advantage of test adaptation is that clear guidelines for the assessment format have also been validated. Decisions regarding using picture stimuli to keep attention, the number of test items to reduce

fatigue effects, and incorporating videos to standardize presentation, have already been determined for test developers.

It is important to clarify the distinction between "translation," defined as a one-to-one transfer without consideration of linguistic differences, and "adaptation," which involves developing a parallel test that "acknowledges the linguistic, cultural, and social conditions of those taking the adapted test while retaining the measurement of the constructs found in the original" (Oakland & Lane, 2004, p. 239). In the examples described above, the new tests were developed to closely resemble the existing tests, but incorporated the specific needs of the target language, therefore, adaptation is the appropriate term to use to describe the process.

<2>New Technologies

As has been mentioned by several test developers in the above descriptions, test delivery through a web-based format is highly conducive to signed language assessment (see Mann & Haug, in press). This format easily incorporates multiple picture/video stimuli and response options, and can also record video data (signing) expressed by test takers. A web-based system allows for multiple people to take the test simultaneously, making the testing process much more efficient. The test results are also recorded and analyzed automatically and can immediately be entered into a database for further comparison and psychometric evaluation. The potential of such a database is that it can also generate reports for test administrators, parents and schools, as is the case for the Sign Language RST online version (Haug, Herman, & Woll, 2014). Another benefit to web-based test formats is the increased mobility and accessibility, as the assessments can be accessed on a variety of different devices. The Sign Language RST online version has

a responsive design that enables test participants to take the test on a tablet or smartphone (Figure 12).

<figure 12 here>

In summary, the lack of standardized, norm-referenced assessment instruments to measure the acquisition of natural signed languages in children has been an enormous gap in both research and education concerning young deaf children and their development. In comparison, standardized assessment measures of speech and language skills of most spoken languages (English, German, French, and so on) are extremely numerous and are constantly being revised and updated (Owens, 2004). Standardized tests of spoken languages allow researchers to communicate more precisely and build on one another's findings with ease. Research in signed language development lags significantly behind with respect to standardized testing. Accurate measurement of signed language skills is particularly critical for children acquiring such languages as their first language, as delays in first language acquisition have a detrimental impact on later learning and literacy development. Valid and reliable measures of signed language development are needed to understand diversity among deaf children in various areas of development and performance, including behavior, social interaction, cognition, spoken language skills, and literacy levels (e.g., Oliva, Lytle, Hopper, & Ostrove, this volume). The assessment tools described in this chapter begin to fill the void and can meet the needs of teachers and researchers in providing appropriate educational programming, monitoring and reporting. More importantly, valid and reliable measures enhance the credibility of signed languages, promote strategies that build on deaf children's visual strengths (Herman & Roy, 2006; Hoffmeister & Caldwell-Harris, 2014; Marshall,

Denmark & Morgan, 2006; Mason, et. al., 2010; McQuarrie & Parrila, 2014), and value sign languages as equal and legitimate languages of instruction in schools.

<2>References

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, 62*(1), 15-20.

Bizer, S., & Karl, A. -K. (2002a). Entwicklung eines Wortschatztests fuer gehoerlose Kinder im Grundschulalter in Gebaerden-, Schrift-, und Lautsprache. Unpublished doctoral dissertation, Fachbereich Erziehungswissenschaften, Universitaet Hamburg.

Bizer, S., & Karl, A. -K. (2002b). Perlesko: Pruefverfahren zur Erfassung lexikalisch-semantischer Kompetenz gehoerloser Kinder im Grundschulalter. Manual. Hamburg: Fachbereich Erziehungswissenschaften, Universitaet Hamburg.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brownell, R. (2000). *Expressive one-word picture vocabulary test*. Novato, CA: Academic Therapy Publications.

Chamberlain, C. & Mayberry, R.I. (2008) American Sign Language syntactic and narrative comprehension in skilled and less skilled readers: Bilingual and bimodal evidence for the linguistic basis of reading. *Applied Psycholinguistics*, 29, 367–388.

Cormier, K., Adam, R., Rowley, K., Woll, B., & Atkinson, J. (2012). *The BSL Sentence Reproduction Test: Exploring age-of-acquisition effects in British deaf adults.* Paper presented at the 34th Annual Meeting of the German Association of

Linguistics, March 7-9, 2012, Frankfurt, Germany.

Dammeyer, J. (2010) Psychosocial development in a Danish population of children with cochlear implants and deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education*, 15 (1), 50-58.

Davidson, K., Lillo-Martin, D. & Chen-Pichler, D. (2014) Spoken English language development in native signing children with cochlear implants. *Journal of Deaf Studies and Deaf Education, 19,* 238-250

DeLana, M., Gentry, M.A., & Andrews, J. (2007). The efficacy of ASL/English bilingual education: Considering public schools. *American Annals of the Deaf, 152*(1), 73-87.

Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research, 41*, 1136–1146.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test (3rd ed.).* Circle Pines, MN: American Guidance Services.

Enns, C., Boudreault, P., Zimmer, K., Broszeit, C., & Goertzen, D. (2014, February). *Assessing children's expressive skills in American Sign Language.* Paper presented at the annual meeting of the Association of College Educators – Deaf/Hard of Hearing, Washington, DC.

Enns, C. & Herman, R. (2011) Adapting the Assessing British Sign Language Development: Receptive Skills Test into American Sign Language. *Journal of Deaf Studies and Deaf Education, 16*, 362-374.

Falkman K.W., Roos, C. & Hjelmquist, E. (2007) Mentalizing skills of non-native, early signers: A longitudinal perspective. *European Journal of Developmental Psychology*, 4(2).

Fosshaug, S. (2010). *Entwicklung der gebärdensprachlichen Kompetenz eines gehörlosen Kindes in einer bilingual geführten Schulklasse: eine Logitudinalstudie.* VUGS Informationsheft Nr. 47. Zürich: VUGS.

French, M. (1999). *Starting with Assessment Toolkit.* Washington, DC: Gallaudet University Press.

Gallon, N., Harris, J., & van der Lely, H. (2007). Nonsense word repetition: An investigation of phonological complexity in children with Grammatical SLI. *Clinical Linguistics and Phonetics, 21*, 435–455.

Gathercole, S. E. (2006). Nonsense word repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543.

Gerner de Garcia, B. (2000). Meeting the needs of Hispano/Latino deaf students. In K. M. Christensen & G. L. Deldago (Eds.), *Deaf plus. A multi-cultural perspective* (pp. 149–198). San Diego, CA: Dawn Sign Press.

Haug, T. (2011). Methodological and theoretical issues in the adaptation of sign language tests: An example from the adaptation of a test to German Sign Language. *Language Testing*, *29*(2), 181–201.

Haug, T. (2011) Approaching sign language test construction: Adaptation of the German Sign Language receptive skills test. *Journal of Deaf Studies and Deaf Education*, 16 (3): 343-361.

Haug, T. (2005). Review of sign language assessment instruments. *Sign Language &*

Linguistics, 8, 61-98.

Haug, T., & Hintermair, M. (2003). Ermittlung des Bedarfs von Gebärdensprachtests für gehörlose Kinder – Ergebnisse einer Pilotstudie. *Das Zeichen*, *64*, 220-229.

Haug, T., Herman, R. & Woll, B. (2014) Constructing an Online Test Framework, Using the Example of a Sign Language Receptive Skills Test. *Deafness & Education International*.

Haug, T., Notter, C., Girard, S., & Audeoud, M. (2015). *Sentence Repetition Test für die Deutschschweizerische Gebärdensprache (DSGS-SRT)*. Unpublished test, Hochschule für Heilpädagogik, Zurich.

Haug, T. & Mann, W. (2008). Adapting tests of sign language assessment for 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., Paludnevičienė, R., Supalla, T., & Bavelier, D. (2008). American Sign Language-Sentence Reproduction 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 9th Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006 (pp. 160-172).

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. (1998). Issues in designing an assessment of British Sign Language Development. *Proceedings of the Conference of the Royal College of Speech & Language Therapists* (pp. 332-337). Liverpool, UK.

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 Standardization of an Assessment of British Sign Language Development for Use with Deaf Children: Final Report, 1998*. Manuscript, Department of Language & Communication Science, City University London, UK.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL Development - Receptive Skills Test*. Coleford, UK: Forest Bookshop.

Herman, R., & Roy, P. (2006). Evidence from the wider use of the BSL Receptive Skills Test. *Deafness and Education International*, *8*(1), 33-47.

Herman, R., Rowley, K., Mason, K. & Morgan, G. (2014) Deficits in narrative abilities in child British Sign Language users with specific language impairment. *International Journal of Language & Communication Disorders.*

Hodge, G., Schembri, A. & Rogers, I. (2014) The Auslan (Australian Sign Language) Production Skills Test: Responding to challenges in the assessment of deaf children's signed language proficiency. Presentation at Disability Studies in Education Conference, July, Melbourne, Australia.

Hoffmeister, R. J. (1994). Metalinguistic skills in deaf children: Knowledge of synonyms and antonyms in ASL. In J. Mann (ed.), *Proceedings of the Post Milan: ASL and English Literacy Conference* (pp. 151-175). Washington, DC: Gallaudet University Press.

Hoffmeister, R. J. (2000). A Piece of the 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: Erlbaum.

Hoffmeister, R. J., & Caldwell-Harris, C. L. (2014). Acquiring English as a second language via print: the task for deaf children. *Cognition*, *132*(2), 229–42.

Hoffmeister, R.J., Caldwell-Harris, C.L., Henner, J., Benedict, R., Fish, S., Rosenburg, P. , Conlin-Luippold, F. & Novogrodsky, R. (2014). The American Sign Language Assessment Instrument (ASLAI): Progress report and preliminary findings. Working paper: Center for the Study of Communication and the Deaf.

Hoffmeister, R.J, Henner, J. & Caldwell-Harris, C.L. (2015) The Psychometric Properties of the American Sign Language Assessment Instrument, Working Paper, Center for the Study of Communication and the Deaf, Boston University

Jackson A. L. (2001) Language facility and theory of mind development in deaf children. *Journal of Deaf Studies and Deaf Education, 6*, 161-176

Jones, A., Herman, R., Botting, N., Marshall, C., Toscano, E. & Morgan, G. (2015) Narrative skills in deaf children: assessing signed and spoken modalities with the same test. Presentation at SRCD Preconference: Development of deaf children, March, Philadelphia, USA.

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.

Kennedy, C.R., McCann, D.C., Campbell, M.J., Law, C.M., Mullee, M., Petrou, S., Watkin, P., Worsfold, S., Yuen, H.M. and Stevenson, J. (2006) Language ability after early detection of permanent childhood hearing impairment. *New England*

Journal of Medicine, 354 (20), 2131-41.

Kubus, O., & Rathmann, C. (2012). *Degrees of difficulty in the L2 acquisition of morphology in German Sign Language.* Paper presented at the 34th Annual Meeting of the German Association of Linguistics, March 7-9, 2012, Frankfurt, Germany.

Labov, W. & Waltesky, J. (1967). Oral Versions of Personal Experiences. In J. Helm (Ed.), *Essays on the verbal and visual arts,* pp. 12-44. Seattle: University of Washington Press.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21*, 202–226.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*, 469–523.

Lillo-Martin, D. (1999). Modality effects and modularity in language acquisition: The acquisition of America Sign Language. In W. Ritchie & T. Bhatia (Eds.), *Handbook of child language acquisition.* San Diego, CA: Academic Press.

MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A., Williams, S. C. R., Suckling, J., Calvert, G. A., & Brammer, M. J. (2002) Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain*, 125(7), 1583-1593.

Mann, W. & Haug, T. (2015). New directions in sign language assessment. In M. Marschark & P.E. Spencer (Eds.), *The Oxford Handbook of Deaf Studies inLanguage*. New York, NY: Oxford University Press.

Mann, W., & Marshall, C. (2012). Investigating deaf children's vocabulary knowledge in

British Sign Language. *Language Learning, 62*(4), 1024–1051.

Mann W., Marshall C. R., Mason K., Morgan G. (2010). The acquisition of sign language: the impact of phonetic complexity on phonology. *Language, Learning, and Devevlopment 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.

Mann, W., Roy, P., & Marshall, C. (2013). A look at the other 90 per cent: Investigating British Sign Language vocabulary knowledge in deaf children from different language learning backgrounds. *Deafness & Education International, 15*(2), 91–116.

Mann, W., Roy, P., & Morgan, G. (2015). Adaptation of a vocabulary test from British Sign Language to American Sign Language. *Language Testing,* 1-20.

Marshall, C. R., Denmark, T., & Morgan, G. (2006). Investigating the underlying causes of SLI: A non-sign repetition test in British Sign Language. *International Journal of Speech-Language Pathology*, *8*(4), 347-355.

Marshall, C., Mason, K., Rowley, K., Herman, R., Atkinson, J., Woll, B., & Morgan, G. (2014). Sentence Repetition in Deaf Children with Specific Language Impairment in British Sign Language. *Language Learning and Development*, 1–15.

Mason K, Rowley K, Marshall C, Atkinson J, Herman R, Woll B & Morgan G (2010) Identifying specific language impairment in deaf children acquiring British Sign Language: Implications for theory and practice. *British Journal of Developmental Psychology*, 28, 33–49.

Mason, T. C. (2005). Cross-cultural instrument translation: Assessment, translation, and statistical application. *American Annals of the Deaf, 151,* 356-370.

McQuarrie, L. & Spady, S. (2012). *The American Sign Language Phonological Awareness Test (ASL-PAT)*, unpublished test. University of Alberta, Edmonton, Canada

McQuarrie, L. & Abbott, M. (2013). Bilingual Deaf students' phonological awareness in ASL and reading skills in English. *Sign Language Studies (Special Issue on Assessment)* 14 (1): 80-100.

McQuarrie, L., & Abbott, M. (2010). Does ASL phonology underlie the ASL-Reading link? Poster presented at the *21st International Congress on the Education of the Deaf (ICED)*. Vancouver, British Columbia.

McQuarrie, L., Abbott, M., & Spady, S., (2012). American Sign Language phonological awareness: Test development and design. *Proceedings of the 10th Annual Hawaii International Conference on Education,* (pp. 1-17), Honolulu, Hawaii.

McQuarrie, L., & Enns, C. (2015). Bridging the gap: Investigating the effects of a signed language phonological awareness intervention on language and literacy outcomes in bilingual deaf children. Paper presented at the *22nd International Congress on the Education of the Deaf (ICED)*. Athens, Greece.

McQuarrie, L. & Parrila, R. (2014). Literacy and linguistic development in bilingual deaf children: Implications of the 'and' for phonological processing. *American Annals of the Deaf (Special Literacy Issue), 159* (4): 372-384.

Menyuk, P., & Brisk, M. E. (2005). *Language development and education: Children with varying language experience.* New York, NY: Palgrave Macmillian.

Meristo, M., Falkman, K. W., Hjelmquist, E., Tedoldi, M., Surian, L. & Siegal, M. (2007). Language and theory of mind reasoning: Evidence from deaf children in bilingual and oralist environments. *Developmental Psychology*, 43, 1156–69.

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.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., & Lee, R. G. (2001). *The syntax of American Sign Language: Functional categories and hierarchical structure*. Cambridge, MA: MIT Press.

Newport, E. & Meier, R. (1985). The acquisition of American Sing Language. In D. I. Slobin (Ed.), The crosslinguistic study of language acquisition, Vol. 1. The Data (pp. 881-938). Hillsdale, NJ: Lawrence Erlbaum Associates.

Norbury, C.F., & Bishop, D.V.M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders*, 38, 287-313.

Novogrodsky, R., Caldwell-Harris, C., Fish, S., & Hoffmeister, R. J. (2014). The Development of Antonym Knowledge in American Sign Language (ASL) and Its Relationship to Reading Comprehension in English. *Language Learning*, *64*(December), 749–770.

Novogrodsky, R., Fish, S., & Hoffmeister, R. (2014). The Acquisition of Synonyms in American Sign Language (ASL): Toward a Further Understanding of the Components of ASL Vocabulary Knowledge. *Sign Language Studies*, *14*(2), 225–249.

Oakland, T. & Lane, H. (2004). Language, reading and readability formulas: Implication for developing and adapting tests. *International Journal of Testing, 4,* 239-252.

Owens, R. (2004). *Language development: An introduction*. Needham Heights, MA: Allyn & Bacon.

Padden, C., & Ramsey, C. (2000). American Sign Language and reading ability in deaf children. In C. Chamberlain, J. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 165-189). Mahwah, NJ: Lawrence Erlbaum Associates.

Prinz, P., & Strong, M. (1998). A Study of the relationship between American Sign Language and English literacy. *Journal of Deaf Studies and Deaf Education, 2*(1), 37-46.

Reese, R., Suggate, S., Long, J. & Schaughency, E. (2010) Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing,* 23 (6), 627-644.

Schembri, A., Wigglesworth, G., Johnston, T., Leigh, G., Adam, R., & Barker, 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. (2003). In Marshcark & Spencer (eds.) *Deaf Studies, Language, and Education*. New York: Oxford University Press.

Sieratzki, J.S., Calvert, G., Brammer, A.M., David, A. & Woll, B. (2001) Accessibility of spoken, written, and sign language in Landau-Kleffner syndrome: a linguistic and functional MRI study. *Epileptic Disorders*, 3(2).

Singleton, J. L., & Supalla, S. J. (2011). Assessing children's proficiency in natural signed languages. In M. Marschark & P.E. Spencer (Eds.), *The Oxford handbook*

of deaf studies, language, and education, volume 1, second edition (pp. 289-302). New York, NY: Oxford University Press.

Snijders, J.T., Tellegen, P.J. & Laros, J.A. (1989) *Snijders-Oomen Non-verbal Intelligence Test. Manual & Research Report.* Groningen: Wolters-Noordhoff.

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: Erlbaum.

Surian, L., Tedoldi, M. & Siegal, M. (2010) Sensitivity to conversational maxims in deaf and hearing children. *Journal of Child Language,* 37, 929–943.

Sutton-Spence, R., & Woll, B. (1999). *The linguistics of British Sign Language: An introduction*. Cambridge, UK: Cambridge University Press.

Tomasuolo, E., Valeri, G., Di Renzo, A., Pasqualetti, P. & Volterra, V. (2012) Deaf children attending different school environments: Sign Language abilities and theory of mind. *Journal of Deaf Studies and Deaf Education*, *18*, 12-29.

Valli, C., & Lucas, C. (1992). *A resource text for ASL users*. Washington, DC: Gallaudet University Press.

Valmaseda, M., Pérez, M. , Herman, R., Ramírez, N & Montero, I. (2013) Evaluación de la competencia gramatical en LSE: Proceso de adaptación del BSL Receptive Skill Test (test de habilidades receptivas). www.cnlse.es/sites/default/files/Evaluacion%20de%20la%20competencia%20gramatical%20en%20LSE.pdf (Accessed 6/3/15)

Vinson, D., 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*, 1079–1087.

Woolfe, T., Want, S.C. & Siegal, M. (2002) Signposts to development: Theory of Mind in Deaf Children. *Child Development,* 73(3), 768-778.

Table 1: Summary of Signed Language Tests

Test/Authors	Purpose	Format	Target Population	Normative Sample
ASLAI Hoffmeister, Caldwell-Harris, Fish, Henner & Rosenburg	Comprehensive measure of receptive: 1) ASL Vocabulary, 2) Reasoning Skills, 3) ASL Syntax, and 4) ASL Text Comprehension	Web-based: - sign to sign - picture to picture - picture to sign - drag & drop - response-only - video event to sign	4 – 18 years: 4 – 7 yrs (7 tasks) 7 – 12 yrs (12 tasks) 12 – 18 yrs (11 tasks)	1. All Deaf children 2. DCDP 3. DCHP
BSL RST Herman, Holmes & Woll	Comprehension of selected aspects of BSL morphosyntax (number/distribution; noun-verb distinction; negation; spatial verbs; handling classifiers; and size-and-shape classifiers)	Video-based (DVD) Web-based version now available	3 – 12 years	135 deaf children (deaf & hearing families); 3 – 13 years; no additional needs; normal non-verbal IQ
ASL-RST Enns, Zimmer, Boudreault, Rabu, & Broszeit	Comprehension of ASL morphology and syntax (number/distribution; noun-verb distinction; negation; spatial verbs; handling classifiers; and size-and-shape classifiers; conditionals; role shift)	Video-based (DVD) Web-based version being developed	3 – 13 years	203 deaf children (deaf & hearing families); 3 – 13 years; no additional needs; normal non-verbal IQ
DGS RST	Comprehension	Web-based	3 – 11 years	54 deaf

Haug	of DGS morphology and syntax (number and distribution, negation, spatial verbs, handling classifiers, and size-and-shape classifiers)			children (deaf & hearing families)
BSL PT Herman, Morgan, Woll & Sutherland	Narrative skills and use of BSL grammar based on a narrative recall task (analysis of narrative content, narrative structure, and BSL grammar)	Video (DVD) elicitation and video recording of narrative recall for later analysis using scoring form	4 – 11 years	71 deaf children (deaf & hearing families); 4 – 11 years; no additional needs; normal non-verbal IQ
ASL PT Enns, Boudreault, & Zimmer	Narrative skills and use of ASL grammar based on a narrative recall task (analysis of narrative content, narrative structure, and ASL grammar)	Video (DVD) elicitation and video recording of narrative recall for later analysis using scoring form Online elicitation, recording, and scoring procedure being developed	4 – 12 years	N/A
BSL NSRT Mann, Marshall, Mason, & Morgan	Ability to repeat nonsense signs of differing phonetic (handshape and movement) complexity in BSL	Video-based Scored for overall correct response, phonological errors, deletion of movements	3 – 11 years	91 deaf children (deaf & hearing families) Hearing control group (n=46)
BSL-VT Mann, Roy, & Marshall	BSL Vocabulary – degree of	Web-based Four tasks: - meaning	4 – 15 years	67 deaf children (deaf & hearing

	strength between form and meaning for core lexicon	recognition - form recognition - meaning recall - form recall		families); 4 – 17 years; variable BSL exposure
ASL-VT Mann	Receptive and expressive ASL vocabulary	Web-based Four tasks: - meaning recognition - form recognition - meaning recall - form recall	6 – 10 years	20 deaf children of deaf parents; 6 – 10 years
ASL-PAT McQuarrie & Spady	Receptive phonological similarity judgment task to assess knowledge of the sublexical properties of sign formation (handshape, location, and movement)	Web-based Video instructions and test items Online scoring of accuracy (number correct), error analysis, and reaction time	4 – 8 years	N/A
DSGS SRT	Sentence repetition task as a global measure of expressive DSGS linguistic skills (phonology, morphology, syntax)	Video-based (DVD) Responses video recorded for scoring (scoring tool being developed)	6 – 12 years	N/A

Table 2: Steps in adapting the BSL RST to DGS (from Haug, 2012)

Steps	Description of steps
1. Review and revision of test stimuli	Picture materials reviewed and changes made, e.g. replacing the red British mailbox with a yellow German mailbox
2. Pilot 1	Suitability of test items established: check for regional variation in three regions with deaf adults and children
3. Adaptation of items	(1) Order of test items (2) Comparability of BSL and DGS linguistic structures (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: (1) Non-signing hearing children and (2) Deaf adults
7. Revisions of first version	Revision of the first version based on Pilot 2: (1) Changes to the pictures (2) Re-filming of items (3) Changes to the layout
8. Planning of main study	(1) Contacting the schools (2) Development and distribution of educational background questionnaires for children
9. Main study	Conducting the main study at five school sites in Germany

List of Figures

---

[i] We would like to thank our colleagues who were kind enough to share their SRT materials for our project: Peter Hauser (USA), Kate Rowley (UK), Christian Rathmann (Germany), Krister Schönström (Sweden), and Pasquale Rinaldi (Italy).

[ii] http://ekids.sgb-fss.ch