Adapting the *Assessing British Sign Language Development: Receptive Skills Test* Into American Sign Language

Charlotte J. Enns*,1, Rosalind C. Herman2

1University of Manitoba

2City University London

Signed languages continue to be a key element of deaf education programs that incorporate a bilingual approach to teaching and learning. In order to monitor the success of bilingual deaf education programs, and in particular to monitor the progress of children acquiring signed language, it is essential to develop an assessment tool of signed language skills. Although researchers have developed some checklists and experimental tests related to American Sign Language (ASL) assessment, at this time a standardized measure of ASL does not exist. There have been tests developed in other signed languages, for example, British Sign Language, that can serve as models in this area. The purpose of this study was to adapt the *Assessing British Sign Language Development: Receptive Skills Test* for use in ASL in order to begin the process of developing a standardized measure of ASL skills. The results suggest that collaboration between researchers in different signed languages can provide a valuable contribution toward filling the gap in the area of signed language assessment.

The education of deaf students has included some significant changes and controversies over the past century. In particular, the use of signed languages in schools has been an issue. Although signed languages were regularly used to educate deaf students at the end of the nineteenth century (Lane, 1984), this changed to an emphasis on oral and auditory skills until about the 1970s (Marschark, Lang, & Albertini, 2002). Over the past few decades, signed languages have again become part of deaf education programs that incorporate a bilingual approach to teaching and learning (Knight & Swanwick, 2002; Mahshie, 1995; Strong & Prinz, 1997).

In order to monitor the success of bilingual deaf education programs, and in particular to monitor the progress of children acquiring signed language, it is essential to develop an assessment tool of signed language skills. Although researchers have developed some checklists and experimental tests related to American Sign Language (ASL) assessment, at this time a standardized measure of ASL does not exist (Singleton & Supalla, 2005, 2011). There have been tests developed in other signed languages, for example, British Sign Language (BSL), that can serve as models in this area. The purpose of this study was to adapt the *Assessing British Sign Language Development: Receptive Skills Test* (*BSL RST*; Herman, Holmes, & Woll, 1999) for use in ASL in order to begin the process of developing a standardized measure of ASL skills.

Background

The key premise upon which all bilingual deaf education programs are based is establishing a first language foundation in a natural signed language. Bilingual programs emphasize first language acquisition in signed language because these languages are considered the most natural and accessible languages for deaf children (Johnson, Liddell, & Erting, 1989; Klima & Bellugi, 1979). Without an established first language, the entire program is brought into question. The primary objective of bilingual deaf education programs is to facilitate the normal acquisition of language, cognition, and social structures through an accessible first language and then build the skills of academic learning and literacy upon this foundation. Therefore, if deaf

*Correspondence should be sent to Charlotte J. Enns, Faculty of Education, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2 (e-mail: ennscj@cc.umanitoba.ca).

students enter school without an established language base, developing their signed language skills must be the focus of education before proceeding with other curricular areas. For this reason, the need for reliable and valid assessments of children's signed language proficiency is essential in furthering the educational objectives of bilingual programs for deaf students.

Determining children's level of signed language proficiency as they begin formal schooling is a major purpose of assessment. The need to monitor children's progress is another purpose of assessment. Deaf children who are having difficulty developing signed 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 so that parents are aware of their child's level of functioning and rate of progress. Accurate assessment can serve a variety of purposes, and this clearly identifies the need for signed language assessment tools. Unfortunately, in the area of signed language acquisition very few commercially available assessment measures exist (Singleton & Supalla, 2005, 2011). As a result, teachers often rely on informal descriptive measures to develop teaching goals and monitor progress (Herman, 1998).

There are legitimate reasons for the paucity of tests in the area of signed language acquisition. Identifying developmental problems in the acquisition of minority languages, whether signed or spoken, is challenging (Johnston, 2004). Norms for these populations often do not exist. There is a lack of controlled elicited data from representative samples of native users of various natural signed languages upon which norms for competency could be established (Schembri et al., 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. Despite these limitations, a variety of signed language assessment measures have been developed and are summarized by Haug on his Web site (www.signlang-assessment.info). Haug lists 15 assessment measures in the area of signed language

acquisition, with particular focus on German Sign Language (DGS)—three tests; Sign Language of the Netherlands (SLN)—three tests; BSL—three tests; and ASL—three tests. Several of the tests listed are adaptations of the "BSL Receptive Skills Test" (into DGS, Italian Sign Language [ISL], Australian Sign Language, and ASL), three tests specifically focus on vocabulary, and another three are observational checklists. Comprehensive communicative assessments are available in DGS (Aachen Test for Basic German Sign Language Competence), SLN (Assessment Instrument for Sign Language of the Netherlands), BSL, and ASL, but none of these tests have been standardized for comparison to a normative population.

Language researchers have defined some key developmental milestones and acquisition patterns in the ASL development of young deaf children (French, 1999; Lillo-Martin, 1999; Newport & Meier, 1985; Schick, 2003). Considerable information is also available regarding the linguistic features of ASL and their relative grammatical complexity, and this can be used to develop guidelines regarding the sequence of acquisition (Neidle, Kegl, MacLaughlin, Bahan, & Lee, 2001; Valli & Lucas, 1992). Although the connection between ASL acquisition research and the development of practical assessment tools needs to be extended, there are several experimental measures that are worthy of review.

The American Sign Language Assessment Instrument (ASLAI; Hoffmeister, 1994, 2000) consists of eight different measures, each developed to assess a particular ASL structure. It provides an in-depth investigation of both language production and comprehension. The ASLAI was used within the framework of a larger research project investigating the relationship of ASL as the first language and English (literacy) as the second language in deaf children. It is not currently available (Haug, 2005).

The Test of ASL (TASL) was also developed as part of a larger study examining the relationship between ASL and English literacy skills (Strong & Prinz, 1997, 2000). The TASL consists of two production measures (Classifier Production Task and Sign Narrative) and four comprehension measures (Story Comprehension, Classifier Comprehension Test, Time Marker Test, and Map Marker Test). The

TASL has been pilot tested and used for research purposes on 155 deaf students aged 8–15 years, but it is not commercially available (Haug, 2005).

The ASL Proficiency Assessment (ASL–PA; Maller, Singleton, Supalla, & Wix, 1999) is a screening tool developed to determine the level of ASL skills of nonnative deaf children. It is based on studies of ASL acquisition that identified the development of eight morphosyntactic structures, including (1) one- and two-sign utterances, (2) nonmanual markers, (3) deictic pointing, (4) referential shifting, (5) verbs of motion, (6) aspects and number, (7) verb agreement, and (8) noun–verb pairs. The test procedure involves eliciting a videotaped language sample through interview with the examiner, peer interaction, and story retelling. The children are scored globally according to three levels of proficiency: Level 3 (16 or more targets produced), Level 2 (11–16 targets), and Level 1 (less than 11 targets). Initial testing involved 80 deaf children, aged 6–12 years, and although some psychometric testing for reliability and validity has been conducted, the measure has not been standardized with large sample norms. As this test is still under development, it is not available to the public (Haug, 2005).

The MacArthur Communicative Development Inventory for ASL (ASL–CDI) (Anderson & Reilly, 2002) is a measure of early vocabulary development in deaf children acquiring ASL. This assessment focuses on infants aged 8–36 months and is completed through parental or caregiver report. The assessment is based on the English version of the CDI (Fenson et al., 1993). Although the ASL–CDI has been shown to be reliable and valid, and it is commercially available, its scope is limited to assessing productive lexical development at the early (8–36 months) preschool level. At the other end of the age spectrum, researchers at the National Science Foundation Science of Learning Center on Visual Language and Visual Learning (VL2), National Technical Institute for the Deaf affiliation, are working to adapt measures of the ASL Sentence Reproduction Test (ASL-SRT) measuring proficiency in adults, for use with children (Hauser, 2008), but these tools are not yet available.

Each of the measures of ASL currently under development has strengths and weaknesses, particularly with regard to availability and purpose of assessment.

Clearly, the need for an assessment tool that can easily, reliably, and efficiently be administered and scored by teachers, as well as be used to monitor progress and provide guidelines for instruction is needed. This need led to the exploration of tests developed in other signed languages that were standardized and commercially available, namely, the *BSL RST.*

The goal of the *BSL RST* (Herman et al., 1999) is to assess understanding of syntactic and morphological aspects of BSL in children aged 3–11 years. The test has both a vocabulary check and a video-based receptive test. The vocabulary check is a simple picture-naming task of 24 items and is used to confirm knowledge of the test vocabulary and to identify any sign variations children may have that differ from those used in the test. If necessary, the test administrator shows the child the test sign and ensures that they can accept this version. If children do not know (unable to name or recognize) more than five of the vocabulary items, the test is discontinued at this point.

The receptive test includes three practice items, followed by 40 test items, organized in order of difficulty and presented by video (tape or DVD format). Test items assess children's knowledge of BSL grammatical structures, including negation, number and distribution, verb morphology, and noun–verb distinction. Administering the test involves the child watching the video of a deaf adult explaining the test procedure and then presenting each test item. There are fade-outs between items that allow the child time to respond. The child responds by pointing to the appropriate picture represented by the signed item from a choice of three or four pictures in a colorful picture booklet. For children who require longer response time, the video can be paused for this purpose. Testing time varies from 12 to 20 min, depending on children's response times. Scoring includes a quantitative raw score (number of items passed) that is converted to a standard score, as well as a qualitative error analysis to describe the pattern of errors made in relation to grammatical structures. The normative data are based on 138 children tested in England, Scotland, and Northern Ireland. The sample included 76 girls, 62 boys; 20 hearing, 118 deaf; 78 from deaf signing families; 23 from established bilingual programs; and

37 from total communication (TC) programs. Through the standardization, it was determined that there was no difference in test performance between hearing and deaf children from deaf families and that there was no difference between children from deaf families and children in established bilingual programs or from children in TC programs with deaf family members who sign. Children in TC programs with no access to BSL outside school performed significantly below the other groups. Representation from all these groups is included within the standardization sample.

The *BSL RST* is the first standardized test of any signed language in the world that has been normed on a population and tested for reliability (Johnston, 2004). For this reason, researchers from several difference countries have chosen to adapt it into other signed languages. The advantage of adapting an existing test rather than developing an original test is that important considerations and decisions have already been evaluated. For example, the *BSL RST* is based on what is 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 et al., 1999). Considering that many signed languages share these important grammatical features, it is likely that test items will be relevant in signed languages other than BSL.

Another important consideration is the composition of the standardization sample given the inconsistent exposure to signed language that occurs for most deaf children. Decisions regarding including hearing children with deaf parents or deaf children with hearing parents attending bilingual or TC programs have already been made and substantiated with research evidence for the *BSL RST*. In addition, clear guidelines for the assessment format have also been validated. These decisions include using pictures versus toys to keep attention but not be distracting; ensuring familiarity with vocabulary through a pretest; keeping items to an appropriate length to avoid excessive memory load; reducing fatigue effects due to length of total test items; and incorporating a video of target structures to standardize presentation and minimize influence by test administrator.

The advantages of adapting an existing test account for the fact that the *BSL RST* has currently been adapted for use in five other signed languages, including French Sign Language, Danish Sign Language, ISL, DGS, and Australian Sign Language (Haug & Mann, 2008). At this point, detailed information regarding the effectiveness of these translated tests is only available for Auslan or Australian Sign Language (Johnston, 2004). It should be noted that the adaptation necessary for using the *BSL RST* in Auslan was minimal as Auslan and BSL can be considered dialects of the same language. Essentially, only two signs included in the test (DOG, PENCIL) are signed differently from BSL. For this reason, the signed stimuli were re-videotaped but the BSL picture book was used and the Auslan scores were compared to the BSL norms. The findings suggested "inflated" scores for the Auslan users when compared to the BSL norms. One reason for this is related to the fact that one test item (WRITE-PENCIL) was not assessing the same construction (noun–verb distinction) in Auslan as it was in BSL; however, the item was kept in the test and it may have inflated some scores. Specifically, five students (of the total 45 students tested) would have reached ceiling if they had not got this item right. Discussion of the adapted test results also considered factors such as age of exposure and length of exposure to Auslan, as well as hearing status. The Auslan sample included hearing children who were participants in a reverse sign bilingualism program and some of these hearing children scored better than the deaf students. This raised issues for the researchers regarding the influence of overall language skills (were the hearing children benefiting more directly from the "mouthing" of English words in conjunction with the signing?) and nonlinguistic factors, such as cognition and age (Johnston, 2004). Clearly, further evaluation is needed regarding the equivalency of the BSL and Auslan versions of the test and whether one set of norms can be applied to both languages. Overall, researchers concluded that early exposure is an important requirement in developing a first language, whether it is signed or spoken (Johnston, 2004).

The interest in adapting existing signed language tests into other languages was examined by Haug & Mann (2008). They begin their discussion by

clarifying 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, as cited in Haug & Mann, 2008, p. 139). If the goal is to develop a test that closely resembles the existing test but incorporates the specific needs of the target language, then adaptation is the appropriate term to use to describe the process.

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 & Mann (2008) through examples involving differences in the categorization of linguistic features (classifiers) between ASL and Swedish Sign Language; lexical differences in ISL and BSL (ISL does not distinguish between "boy" and "child" but BSL does); and morphosyntactic issues, such as more devices for negation (ISL) or less variety of devices for negation (French Sign Language) compared to BSL. 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 mailbox that is in contrast to a 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. This article has worked within such a framework and therefore provides valuable insights into the similarities and differences between assessing the receptive skills of children learning BSL and children learning ASL. Some of these differences were easily resolved through the modification of test stimuli, but others required more significant changes to the test. The study also reinforces the benefit of collaboration among researchers in advancing better understandings of natural signed language acquisition and measurement.

Method

The adaptation of the *BSL RST* into ASL would not have been possible without the permission and guidance of the *BSL RST* authors. According to these authors, the process of adapting the *BSL RST* into ASL included the following phases:

1. Consultation with ASL linguists and adult native ASL signers to determine the following:
   (a) the suitability of the test vocabulary, in particular, the presence of regional alternatives,
   (b) the suitability of a direct translation of each existing BSL sentence into ASL,
   (c) whether the existing distracter pictures were viable alternatives for ASL users,
   (d) the need to add test sentences to reflect ASL linguistic constructions that were not currently included and to replace current BSL structures not represented in ASL.

2. Development of new test items identified in 1(d) above (note: this required developing more items than necessary in case not all proved to be equally effective).

3. Redrawing of any culturally inappropriate images (e.g., mail box, steering wheels on right side, etc.).

4. Recording of new test video in ASL to include the test instructions in a child-friendly register.

5. Piloting of translated sentences and new sentences on a sample of typically developing native signers within the recommended age range 3–11 years to determine:
   (a) effectiveness of test items and
   (b) developmental order of difficulty.

6. Item analysis to determine:
   (a) any items that are too easy (passed by all) or too hard (failed by all),
   (b) ability to discriminate based on age,
   (c) developmental order of difficulty.

7. Standardization on a larger sample to develop norms by age.

This article describes the implementation of the first six phases of the test adaptation process as the seventh phase, standardization on a larger sample, has not been completed at this time.

The first phase was completed by assembling a panel of ASL consultants consisting of two university researchers (both hearing; one in Education and the other in ASL linguistics) and five teachers (all deaf and native ASL users; two college instructors, two school teachers, and one early childhood educator). Over a series of sessions, the consultants reviewed each of the *BSL RST* test items and determined whether they were suitable for use in ASL. The results indicated that 28 of the 40 test items did not require changes—direct translation into ASL was possible and would appropriately assess the parallel ASL grammatical structure. For example, negation is assessed in Item 3 of the *BSL RST* with the stimulus sentence "ICE-CREAM NOTHING" (BSL gloss) and response pictures of a boy with a single cone, a boy with a double cone, a girl with a cone, and a boy without any ice cream. This item translated into ASL, and using the same response pictures would effectively assess children's understanding of the negative term "NOTHING" in ASL. Similarly, Item 10 of the *BSL RST*, "TWO-PEOPLE-MEET" (BSL gloss—signed with index finger classifiers on each hand moving toward each other), was designed to measure spatial verb morphology and achieved the same goal in ASL.

Of the 12 remaining test items that did require modification, eight items required changes to target sentences (and development of new items) and four items required changes only to the pictures. In addition, the panel recommended adding a fourth picture to the six *BSL RST* items that only had three picture response choices in order to keep all test items consistent with four choices. Some of the test items that required modification were due to content. For example, an item in the *BSL RST* assessing role shift involved a boy hitting a girl, and this was modified to the more socially appropriate behavior of the boy tapping (for attention) the girl instead. The item still measured children's understanding of role shift in ASL. Similarly, the noun–verb distinction item in the *BSL RST* using the signs for "PENCIL" and "WRITE" (BSL

glosses) was not appropriate in ASL as the signs are different and was therefore replaced with "CHAIR" and "SIT" (ASL glosses). The modifications required to the pictures included changing the steering wheels to the left side of the vehicles and altering mailboxes and train logos.

In Phase 3, the redrawing of test items, it was decided that all the test pictures should be redrawn to ensure that the style of the pictures was consistent throughout the test. The drawings were digitized and printed into a book format. Phase 4 involved recording the ASL test sentences on video. This was completed in an appropriate studio space using high-quality video-recording equipment. The person signing the ASL sentences was a deaf teacher, with deaf parents, and a very fluent ASL user. She was also involved in the adaptation process, so was familiar with the test items and testing procedure. She was able to present the target sentences very clearly and in a child-friendly manner.

The first adapted version of the ASL Receptive Skills Test (ASL RST) was a similar format to the *BSL RST* in that the children watched video-recorded ASL sentences and selected pictures to match. In the ASL version, the child always selected from a choice of four pictures, whereas in the *BSL RST*, six items only had a choice of three pictures. The adapted test included a vocabulary check of 20 items (two less than the *BSL RST*), 3 practice items (same as the *BSL RST*), and 41 test items (one more than the *BSL RST*). Vocabulary and signs included in the adapted ASL test were selected carefully to minimize possible regional sign variations and make the test applicable to children throughout North America. This was a lesson learned from the authors of the *BSL RST* as regional variation had a significant impact on their test, which was not the case in ASL. The adapted ASL test also assessed the same grammatical categories as the *BSL RST*:

1. Number/distribution (including spatial arrangements of objects, e.g., a row of parked cars)
2. Negation (including head shake with signs or negative signs, e.g., not, never, nothing)
3. Noun/verb distinctions (including similar signs with different movements to distinguish object

vs. action, e.g., single movement/hold for "sit" vs. small repeated movement for "chair")

4. Spatial verbs (including the use of classifiers to depict location, e.g., a car behind a house, and classifiers depicting action, e.g., a bicycle going over a hill)

5. Size/shape specifiers (including classifiers to show the attributes of people and objects, e.g., thin stripes on a shirt)

6. Handling classifiers (including classifiers to indicate how objects are held, e.g., eating a sandwich).

It was determined that the children participating in the pilot testing should be native ASL users or more specifically that they be deaf and have deaf parents who have exposed them to ASL from birth. The purpose for limiting the pilot sample in this way was to ensure that the test reflected the appropriate developmental sequence of ASL acquisition based on fully accessible exposure to the language. Given the size of typical Canadian Deaf communities, it was not possible to recruit enough children in one location who met this criterion. For this reason, testing was conducted in Manitoba, Alberta, Ontario, and Minnesota. Ethics approval for pilot testing was obtained through the appropriate university and school boards. Families were recruited through the schools and written informed consent was obtained from all parents, and written consent or assent in ASL (depending on age) was gained from all children prior to testing. All testing was administered by a deaf research assistant (native ASL user) and videotaped to confirm the accuracy of scoring.

Two rounds of Phase 5 (pilot testing) and Phase 6 (item analysis) were necessary to finalize the adapted ASL version of the *BSL RST*. The results of the first round of pilot testing revealed a significant correlation between children's age and raw score; however, this correlation was not as high as the children got older. Essentially, the older children reached the ceiling of the adapted ASL test, and therefore, it did not distinguish children's receptive skills beyond the age of 10 years. Further analysis of test items and ongoing consultation with the *BSL RST* primary author indicated that modifications to stimulus sentences and distracter

pictures were necessary in order to more effectively assess the intended ASL grammatical features. The specific procedures and results of the two rounds of pilot testing and item analysis are discussed in the next section.

Results

Pilot Testing 1

The first round of pilot testing included 47 children, from both Canada and the United States, between the ages of 4 and 13 years. The 47 children were all deaf with deaf parents and no identified disabilities or learning difficulties and included 27 males and 20 females. The test was easy to administer and took approximately 15 min or less for each student to complete. In the first round of pilot testing, all the children attempted all 41 test items regardless of the number of incorrect responses as it was not known whether following the order of difficulty in BSL would be similar in ASL. Similar to the administration of the *BSL RST*, no repetition of items was allowed except with the youngest children (4 years of age). The interscorer reliability was very high; all 47 tests were rescored and only 2 differed by one point each. This is not surprising given the nature of the test and that the task of the scorer is simply to mark the child's response on the score sheet. The relationship between age and raw score was compared using a Pearson correlation coefficient and revealed a strong correlation ($r = .711$, $p < .001$). The correlation between age and raw score was also analyzed with only the children 6 years of age and older ($n = 40$), and although this correlation was still significant, the r value was lower ($r = .589$, $p < .001$). The results of the first round of pilot testing are displayed in Figure 1a and b.

Item analysis indicated that five test items were passed by all children; however, it should be noted that no data were collected with 3-year-olds, and perhaps these items would have been more challenging for that age group. It was also noted that no items were failed by all children. These results contributed to the high scores for children of 8 years and older and the limited ability of the test to distinguish ASL abilities of children over 10 years of age. Please refer to Table 1 for a summary of the first round of pilot test results.
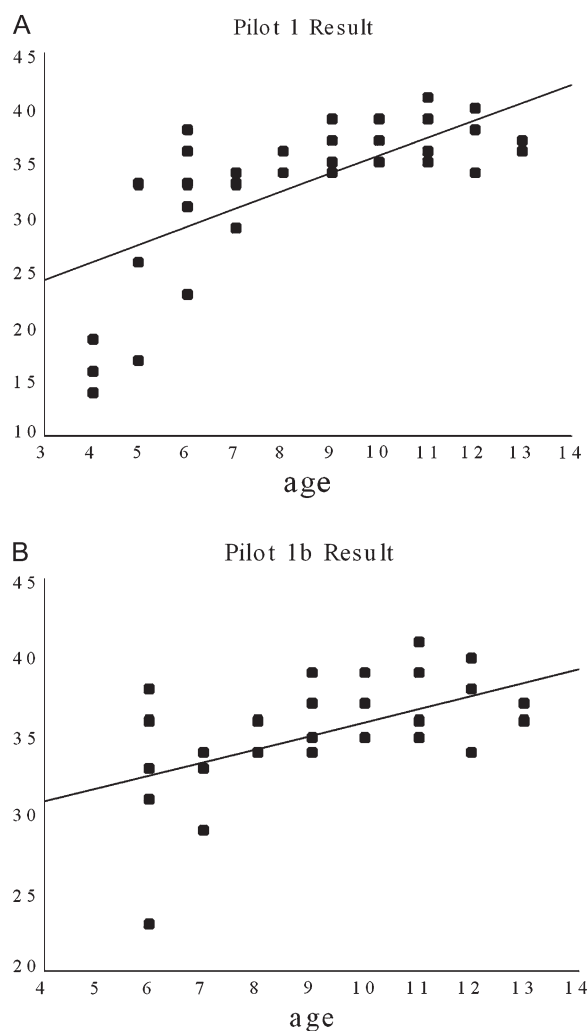
Figure 1 (a) Pilot test 1—all children ($n = 47$). (b) Pilot test 1–children 6 years and older ($n = 40$).

A more thorough analysis of error patterns and consultation with signed language test developers revealed ways to modify test items to make them more challenging and effective in assessing the target ASL structures. Revisions were needed for 23 of the original 41 test items, including changes to distracter drawings (11 items), signed stimulus sentences (4 items), and changes to both drawings and signed sentences (8 items). The revisions made to the distracter pictures ensured that the children needed to use their understanding of ASL grammar to answer the questions rather than being able to deduce the correct answer simply by eliminating irrelevant pictures. Although the pictures were related and therefore effective distracters in BSL, for the ASL version it was

Table 1   First-round pilot testing results

Age (years)	No. of children	Mean raw score (max 41) and range
4	3	16.3 (14–19)
5	4	27.3 (17–33)
6	6	32.3 (23–36)
7	5	32.6 (29–34)
8	3	34.7 (34–36)
9	7	35.6 (34–39)
10	5	37.0 (35–39)
11	5	37.2 (35–41)
12	5	38.0 (34–40)
13	4	36.8 (36–37)

necessary to replace some pictures with items that resembled the target ASL sign more closely than what was used in the *BSL RST*. For example, when the target was an open book lying on the bed (ASL CL: B palm up), the *BSL RST* distracter item, a brush, was replaced with a shirt to more closely resemble the ASL CL: B handshape but with the palm down. Similarly, for the target two rows of beds (ASL CL: N straight fingers), distracter pictures were changed from a bunk bed and a single bed to two rows of chairs (ASL CL: N bent fingers) and two rows of pencils (ASL CL: IN-DEX) to again more closely resemble the ASL target structure. Revisions were also made to some of the signed stimulus sentences. These included shortening the hold or length of the action on noun–verb distinction items, for example "DRINKING" and "DRIV-ING," and ensuring that the signers' hands returned to a neutral position at the end of each item. In some cases, it was necessary to revise both the pictures and the signed stimulus. For example, the ASL item "EAT BIG-SANDWICH" (represented as a submarine-type sandwich held with one hand above the other), involved replacing pictures of eating an apple and eating chips, with eating a watermelon and eating a hamburger—foods that more closely resembled holding a sandwich in ASL. In addition, the handling classifier needed to be signed more clearly in the stimulus sentence. The ASL item, "(NO)-REACH," was modified so that the sign was not held as long, and one of the distracter pictures was modified to increase the reach of a boy struggling to climb onto a chair, where previously his action was not as similar as the target item of the boy reaching for a teddy bear.

In addition to the modifications made to existing test items, four new items were added to assess understanding of role shift (where the speaker embodies two characters marked by shoulder shift and eye gaze, e.g., the mother giving the apple and shifting to the child accepting the apple) and conditional clauses (including the use of raised eyebrows and the ASL sign "SUPPOSE" to mark the first clause and indicate that the second event/clause is dependent on the first). These constructions are considered to reflect more complex ASL grammar (Emmorey & Reilly, 1995), and the intention was to add items that the older children would find more challenging. Clearly, the addition of items involving new grammatical structures is a significant departure from direct test adaptation; therefore, the decision to include these items is discussed further in the next section. The final modification to the adapted ASL test was to reorder the items to more accurately reflect the developmental level of difficulty according to the number of children who passed each item.

Pilot Testing 2

A second round of pilot testing was required for the revised version of the adapted ASL test. The revised version was a similar format to the initial adapted ASL test in that the children were required to watch the video-recorded ASL sentences and then select the appropriate picture from a choice of four. It also included a vocabulary check of 20 words and 3 practice items, but the total test items were increased from 41 to 45 test items. The grammatical categories assessed through these items included the six previous structures (number/distribution, negation, noun/verb, spatial verbs, size/shape specifiers, handling classifiers), as well as the two additional categories of role shift and conditionals.

The second round of pilot testing included 34 children (18 males and 16 females) between the ages of 3 and 13 years. These children were again recruited from the same schools in Minnesota and Ontario and therefore included 29 of the children from the first round of pilot testing but tested 1 year later. The sample again represented all deaf children with deaf parents and no identified disabilities or learning issues. The results of the retesting indicated that the modifications made to previous test items and the new test items had effectively made the test more challenging and more clearly distinguished children's skills at different ages. Specifically, the analysis comparing age and raw score showed a significant correlation and high r value when all 34 children were included ($r = .821$, $p < .001$), as well as when only the children 6 years and older ($n = 20$) were included ($r = .719$, $p < .001$). These data are represented in Figure 2a and b. It should be noted that in this round of pilot testing, the test was discontinued when children obtained five consecutive incorrect responses; therefore, not all children attempted all 45 test items. The test was discontinued for 13 of the 34 children, who were all under the age of 8 years. A column of the average number of attempted items for each age group has been added to Table 2 to reflect the pilot testing data more accurately. Please refer to Table 2 for more specific results of the second round of pilot testing.

Table 2 Second-round pilot testing results

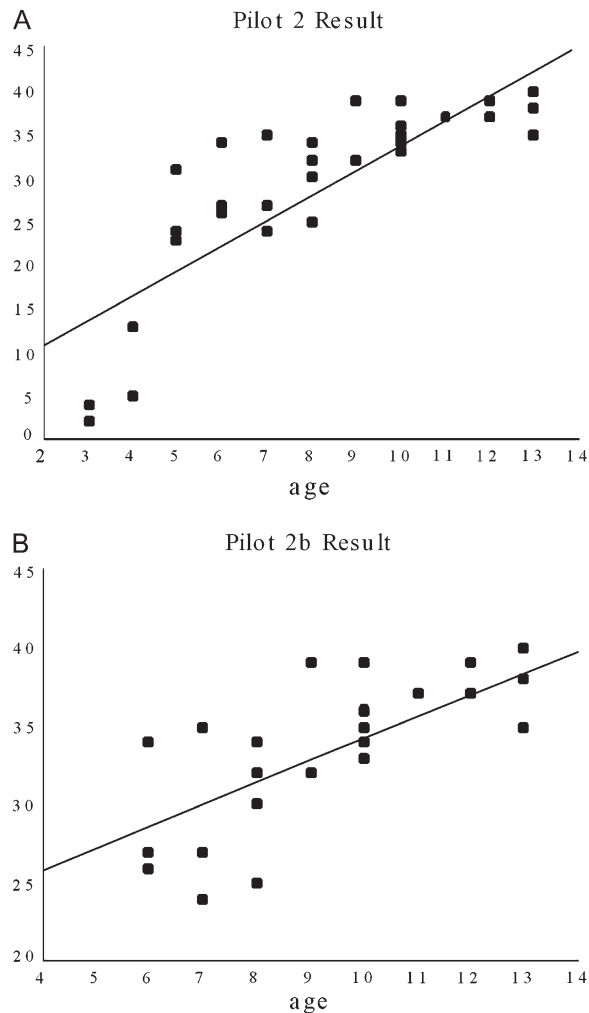| Age (years)	No. of children	Mean raw score (max 45) and range	Mean attempted test items
3	2	3.0 (2–4)	12
4	3	7.7 (5–13)	22
5	3	26.0 (23–31)	36
6	3	29.0 (26–34)	42
7	3	28.7 (24–35)	38
8	4	31.5 (30–34)	45
9	2	35.5 (32–39)	45
10	6	35.6 (33–39)	45
11	1	37.0 (37)	45
12	2	38.0 (37–39)	45
13	5	37.2 (35–40)	45

Figure 2   (a) Pilot test 2—all children (*n* = 34). (b) Pilot test 2–children 6 years and older (*n* = 20).

Discussion

A finalized version of the ASL RST was developed based on the results from the second round of pilot testing. Modifications included deleting three test items and reordering test items to more appropriately reflect the developmental sequence of language acquisition, as indicated by the number of children who correctly answered each test item. The deleted items were considered redundant (several other items were measuring the same grammatical structures) or, in one case, culturally inappropriate (the item included an escalator and children from rural areas were often not familiar with this concept). In its finalized version, the ASL RST includes a vocabulary check

of 20 words, 3 practice items, and a total of 42 test items. Eight grammatical categories are assessed through these items, including number/distribution, negation, noun/verb distinction, spatial verbs (location and movement), size/shape specifiers, handling classifiers, role shift, and conditionals.

The primary goal of this study was to end up with an assessment measure in ASL that functioned in equivalent fashion to the *BSL RST*. That goal was accomplished, but the study has also contributed to a much greater appreciation for the process of test adaptation. In retrospect, given the revisions that were required for numerous test items and distracter pictures, along with the addition of new test items, the value of adapting an existing test over developing a new one must be reconsidered.

One advantage to adapting an existing test was that the format and guidelines for implementation were already determined. It certainly was beneficial to base the ASL RST on the BSL template in establishing a workable test format. Features such as the vocabulary pretest ensured that the skills being assessed were related to children's comprehension of grammatical structures, not their knowledge of lexical items. Similarly, the use of video to consistently present the ASL stimuli was an excellent feature incorporated from the BSL version. The length of test items and the overall length of the test (number of items) also aligned closely with the BSL version to ensure that memory and fatigue were not factors influencing test results. The format of the ASL RST did differ from the BSL version in two ways. The first was that all test items included a choice of four possible responses to provide consistency and equalize the chances of children guessing the correct answer. In the *BSL RST*, 6 of the total 40 test items had only three picture responses. The second change in format to the ASL RST was that in the final version, the video (DVD) included both the ASL stimuli and the picture responses. The advantage to having both the stimuli and the responses on the same screen was that children did not need to shift their eye gaze from screen to picture book, thereby reducing the chances of being distracted from the task or missing parts of the signed stimulus. It was still possible to modify the video presentation to the needs of each child by pausing to allow

longer viewing of the picture responses if required. In the future, a computer-based version of the test may be developed to build on this same-screen advantage, but also improve scoring accuracy and accessibility.

Another advantage to adapting an existing test was that significant grammatical features and their relationship to language proficiency were identified. Although it was certainly true that the six grammatical categories featured in the *BSL RST* were also the primary focus of the items in the ASL RST, careful consideration of linguistic differences between the two languages was required in adapting and developing specific test items. This was particularly evident in developing appropriate distracter pictures. The translation of signed language stimulus phrases or sentences from BSL to ASL was generally straightforward due to the similarities in the way the languages used verb morphology and space to express grammatical concepts (verbs of motion/location, size and shape specifiers). Therefore, it was initially assumed that the picture responses would not require any modification. After the first round of pilot testing, it was revealed that distracter items that were effective in BSL were not necessarily effective in ASL. For example, a number/distribution test item in BSL included response pictures of rows of parked cars as well as books on a shelf. In BSL, both books and cars are designated with a flat hand ("B" handshape in ASL); however, in ASL the designations differ significantly ("B" handshape for books; "3" handshape for cars). Keeping the picture of books on a shelf as a distracter item in the ASL test was not effective as the structure it represented was not similar enough to the target structure. Our experience emphasized that in the process of test adaptation, it was just as important to consider the appropriateness of response items (in this case target and distracter pictures) as it was to consider the stimulus items. We caution researchers in using the same picture responses across signed languages.

Another consideration that was identified through the pilot testing process was the need to add items involving more complex grammatical structures so that the test would distinguish skills in children beyond the age of 10 years. The addition of conditionals and role shift structures certainly deviated from strictly adapting the existing test as multi-clausal items were not part of the *BSL RST*. However, the need for including more complex items was acknowledged by the authors of the *BSL RST* and was an area of consideration for future revisions of the test (Herman, R. C., personal communication, May 2009). This is an example of how collaboration among researchers involved in developing and adapting assessment measures can advance the understanding of acquisition in both languages. The work of developing effective measures of signed language acquisition contributes to understanding signed language grammatical systems and patterns of sociolinguistic variation. Inconsistencies and gaps continue to exist in the knowledge of signed language acquisition, even for well-documented languages like ASL; therefore, descriptive investigation, such as in this article, contributes to our confidence in designing valid and reliable measures of signed language grammatical proficiency (Schembri et al., 2002).

Despite the fact that numerous revisions were required to many of the test items (either to adjust for cultural and linguistic differences or improve the accuracy of the translation) and that new grammatical categories had to be added, the test adaptation process was considered a worthwhile endeavor. In the end, 30 items in the ASL RST very closely resembled those of the original 40 *BSL RST* items, with 23 being identical translations and 7 items having similar signed stimuli and only slight differences in the response pictures.

The finalized version of the ASL RST now requires standardization. This will involve testing a minimum of 20 children at each age level between 3 and 12 years (total of 200 children) to ensure that psychometric measures can be applied and have significant statistical power. Once these data have been collected, the test will provide users with a standard score for what is expected at each age level and it can, therefore, be determined if children are developing an understanding of ASL age appropriately. A standardized measure is beneficial in determining children's proficiency as well as monitoring and reporting on their progress as they acquire ASL. It is anticipated that the ASL RST will also be helpful in identifying the strengths and weaknesses of atypical ASL learners. The *BSL RST* has already proven to be useful in this

way (Mason et al., 2010), as well as with children beyond the age of 12 years due to the numerous deaf children who are late learners of ASL and only have the opportunity to learn signed language when they enter school.

In conclusion, the project to adapt the *BSL RST* for use in ASL has contributed to both practical and research purposes. It provides teachers with a useful and manageable assessment tool to help them deliver appropriate educational programming, as well as to monitor and report on progress in signed language learning. This article also contributes to filling the gap in research regarding the credibility of ASL as a language of instruction in schools. Finally, the process of test adaptation provides insights regarding the similarities and differences between signed languages and through research collaboration facilitates a better understanding of the process of signed language acquisition.

Funding

Conflicts of Interest

No conflicts of interest were reported.

Acknowledgements

References

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education, 7,* 83–119. doi:10.1093/deafed/7.2.83.

Emmorey, K., & Reilly, J. (Eds.). (1995). *Language, gesture, and space*. Hillsdale, NJ: Lawrence Erlbaum.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., & Reilly, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

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

Haug, T. (2005). Review of sign language assessment instruments. *Sign Language & Linguistics, 8,* 61–98. doi:10.1075/sll.8.1.04hau.

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,* 138–147. doi:10.1093/deafed/enm027.

Hauser, P. (2008). *Deaf studies laboratory. VL2 Vision Newsletter, 1 (Spring)* (pp. 5). Washington, DC: Gallaudet University.

Herman, R. (1998). The need for an assessment of deaf children's signing skills. *Deafness and Education, 22,* 3–8.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing British Sign Language development: Receptive skills test*. Gloucestershire, UK: Forest Bookshop.

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: Lawrence Erlbaum Publishers.

Johnson, R. E., Liddell, S., & Erting, C. J. (1989). *Unlocking the curriculum: Principles for achieving access in deaf education (Working Paper 89-3)*. Washington, DC: Gallaudet University, Gallaudet Research Institute.

Johnston, T. (2004). The assessment and achievement of proficiency in a native sign language within a sign bilingual program: The pilot Auslan receptive skills test. *Deafness and Education International, 6,* 57–81. doi:10.1002/dei.168.

Klima, E., & Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.

Knight, P., & Swanwick, R. (2002). *Working with deaf pupils: Sign bilingual policy into practice*. London, UK: David Fulton Publishers.

Lane, H. (1984). *When the mind hears: A history of the deaf*. New York, NY: Random House.

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.

Mahshie, S. N. (1995). *Educating deaf children bilingually*. Washington, DC.: Gallaudet University Press.

Maller, S. J., Singleton, J. L., Supalla, S. J., & Wix, T. (1999). The development and psychometric properties of the American Sign Language Proficiency Assessment (ASL-PA). *Journal of Deaf Studies and Deaf Education, 4,* 249–269. doi:10.1093/deafed/4.4.249.

Marschark, M., Lang, H., & Albertini, J. (2002). *Educating deaf students: From research to practice*. New York, NY: Oxford University Press.

Mason, K., Rowley, K., Marshall, C., Atkinson, J. R., 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. doi:10.1348/026151009X484190.

Mason, T. C. (2005). Cross-cultural instrument translation: Assessment, translation, and statistical application. *American Annals of the Deaf, 150,* 67–72. doi:10.1353/aad.2005.0020.

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. doi:10.1353/sls.2004.0005.

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.

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,* 18–40. doi:10.1093/deafed/7.1.18.

Schick, B. (2003). The development of American Sign Language and manually coded English systems. In M. Marschark & P. Spencer (Eds.), *Deaf studies, language, and education* (pp. 219–231). New York, NY: Oxford University Press.

Singleton, J. L., & Supalla, S. J. (2005). Assessing children's proficiency of natural signed languages. In M. Marschark & P. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (pp. 289–302). Oxford, UK: Oxford University Press.

Singleton, J. L., & Supalla, S. J. (2011). Assessing children's proficiency of natural signed languages. In M. Marschark & P. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (Vol. 1, 2nd Ed, pp. 306–321). New York, NY: Oxford University Press.

Strong, M., & Prinz, P. (1997). A study of the relationship between American Sign Language and English literacy. *Journal of Deaf Studies and Deaf Education, 2,* 37–46. doi:10.1093/deafed/2.1.37.

Strong, M., & Prinz, P. (2000). Is American Sign Language skill related to English literacy? In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 131–142). Mahwah, NJ: Lawrence Erlbaum Publishers.

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