End Note

Adapting Tests of Sign Language Assessment for Other Sign Languages—A Review of Linguistic, Cultural, and Psychometric Problems

Tobias Haug
University of Applied Sciences for Special Needs (HfH), Zurich, and
Hamburg University

Wolfgang Mann
Deafness, Cognition and Language Research Centre and
City University, London

Given the current lack of appropriate assessment tools for measuring deaf children's sign language skills, many test developers have used existing tests of other sign languages as templates to measure the sign language used by deaf people in their country. This article discusses factors that may influence the adaptation of assessment tests from one natural sign language to another. Two tests which have been adapted for several other sign languages are focused upon: the Test for American Sign Language and the British Sign Language Receptive Skills Test. A brief description is given of each test as well as insights from ongoing adaptations of these tests for other sign languages. The problems reported in these adaptations were found to be grounded in linguistic and cultural differences, which need to be considered for future test adaptations. Other reported shortcomings of test adaptation are related to the question of how well psychometric measures transfer from one instrument to another.

The development and availability of appropriate test instruments for sign languages is of practical as well as theoretical significance. Well-made tests can document the effects at many levels of the language of different ages of initial exposure to the language as well as the effects on language fluency depending on whether the language is as a first or second language (Mayberry & Eichen, 1991; Mayberry, Lock, & Kazmi, 2002; Morford & Mayberry, 2000).

Currently, only a small number of tests have been developed to assess natural sign languages[1] (Haug, 2005), and most of these tests are still works in progress. There is even less empirical data to document the need for such tools, with few exceptions (e.g., Haug & Hintermair, 2003; Herman, 1998; Mann & Prinz, 2006). Furthermore, although the use of standardized tests seemed to be well perceived in some countries, this perception may not be equally shared in other countries or tests are not consistently used throughout the country (e.g., in Sweden; Schönström, Simper-Allen, & Svartholm, 2003).

For these reasons, researchers looking for a tool to assess the natural sign language of their country frequently turn to existing tests developed for another natural sign language as a template. However, the attempt to measure similar, or identical, constructs across different languages by adapting or translating tests often results in complications. Among the most common factors influencing successful test adaptation across languages and modalities are differences in linguistic structures and cultural influences. Other issues that require great caution include the adaptation of established psychometric properties of the source test for a new version of the test measuring a different sign language.

Although such complications have often been pointed out in studies on spoken languages (e.g., Rapp & Allalouf, 2003; Solano-Flores, Trumbull, & Nelson-Barber, 2002), little documentation exists on the nature and effects of these complications

for sign languages (Mason, 2005). Therefore, the main objective of this review is to raise and discuss issues related to test development as they specifically apply to the assessment of natural sign languages and the adaptation of tests from one sign language to another.

We will base our discussion of test-related problems and other issues on two existing sign language tests, the Test of American Sign Language (TASL; Prinz, Strong, & Kunze, 1995) and the British Sign Language (BSL) Receptive Skills Test (Herman, Holmes, & Woll, 1999). As both of these instruments have been adapted for a number of other natural sign languages, they provide excellent examples for illustrating the possible complications that may influence the success of such an undertaking.

Translation Versus Adaptation

It is important at the offset to make a distinction between two terms that are often used interchangeably in the literature on test development: test "translation" and test "adaptation".

The term "translation"—even if it may not always be used that way—technically refers to a one-to-one transfer without any consideration of linguistic differences. A translated test should not include any kind of target language substitutions for items in the source test that, despite their linguistic significance for the target language, are not part of the source language. For example, a test to assess language proficiency in spoken English would be unlikely to include any items related to gender as this grammatical feature rarely occurs in English. In a version of such an English test for German, it would thus be difficult to test for gender, despite the fact that in German, gender represents an important grammatical category. As a result, a translated test may provide only limited assessment of grammatical development in the target language.

Geisinger (1994) uses the term "adaptation" rather than "translation" when referring to the transfer of a test from one natural language to another one. Adaptation takes into account both linguistic and cultural differences and involves more flexibility in test construction. The following definition by Oakland and Lane (2004, p. 239) illustrates the many facets that are inherent in the adaptation process:

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

For the remaining part of this review, we will use the term "adaptation" as it incorporates the notion of developing a test for the target language which remains as close as possible to the source language while, at the same time, continues to meet the specific needs of the target language.

Before taking a closer look at the two test instruments that were adapted for other natural sign languages, a brief description of psychometric issues is provided in the next section.

Psychometric Issues

Test developers need to provide evidence for the effectiveness of their instrument based on appropriate psychometric measures. Although those types of measures on test construction and development, which have been reported in the literature, show variation (e.g., Kline, 2000), they all serve the purpose of evaluating a test instrument and/or providing information on participants' test behavior. The measures most commonly applied to describe how test takers' behavior relates to the evaluation of their performance are reliability, validity, and standardization.

Reliability

Reliability can be measured in a number of ways, although there are two types of evidence that are most commonly reported on by researchers. One refers to stability over time, the second to internal consistency. The reliability of a test over time is known as "test–retest reliability" (Kline, 2000, p. 7) for which subject scores that were obtained on two different occasions are correlated. The higher the correlation, the more

reliable is the test. A minimum of .8 in a test–retest should be reached to show a correlation (Kline, 2000, p. 11). The "internal consistency" of a test refers to "the degree to which scores on individual items or group of items on a test correlate with one another" (Davies et al., 1999, p. 86). A measure of internal consistency includes statistical procedures such as the split-half analysis (e.g., using different sets of test items).

"Interrater reliability" refers to the level of agreement between two or more raters on a test taker's performance (Davies et al., 1999, p. 88), for example, to compare the scoring of certain grammatical features that a deaf child performed on a production task that has been videotaped and then rated by two different raters and then compared.

Validity

The main claim for a valid test is that it really measures what it claims to measure (Kline, 2000). With regard to deaf test taker, this could mean whether an assessment of sign language vocabulary really measures the vocabulary knowledge in deaf children or not. There are several types of validity, for example, item or content validity, concurrent validity, predictive validity, and construct validity. Each of these types of validity requires different evidence.

One of the prerequisites for assuring "item" or "content validity" in a test of sign language skills is the close collaboration with deaf native signers during the developmental stage (Singleton & Supalla, 2003, p. 297). "Concurrent validity" can be shown by a high correlation between the targeted test and another test that measures the same variable or construct. However, given the very small number of other sign language tests, this kind of comparative psychometric measure is difficult to carry out. An example of "predictive validity" would be the high correlation between results of a sign language proficiency test and the results of a standardized literacy test, which indicated that sign language proficiency is a predictor for literacy skills. "Construct validity" of a language test provides an indication to what extent the test instrument represents the theory of language learning that serves as underlying construct (Davies et al., 1999).

Only a few tests for American Sign Language (ASL, or other sign languages) have any measures of reliability and validity compared to tests for spoken English, such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), making this one of the major drawbacks for current sign language research.

Standardization

An additional issue that can affect the psychometrics of a test is its process of standardization. The success of this process depends on several variables including

- the size of the population that the sample represents (here, the population of deaf children) and
- the homogeneity (or heterogeneity) of the population (Kline, 2000, p. 51; e.g., differences in parents' hearing status and the diverse cultural and linguistic background).

For developers of sign language tests, this leads to the following questions: (a) What is considered the ideal sample size that represents the entire population? and (b) What is the reference group for which the test will be standardized, taking into consideration the heterogeneity within the deaf population?

In sum, the detailed documentation of the psychometric properties used for the development of the test as well as for the psychometric characteristics for each source or target sign language remain key elements in successfully determining the extent to which different sign language test versions measure the same underlying construct. Such documentation needs to be presented in a format that facilitates the standardization of the instrument. Currently, most of the originally developed tests to measure deaf students' skills in a natural sign language (e.g., ASL, French Sign Language [LSF]) do not meet such requirements.

Descriptions of Two Sign Language Tests

In this section, we will give a brief description of the structure and psychometric data of the TASL and the BSL Receptive Skills Test.

Test of American Sign Language

The TASL was developed within the framework of a larger cooperative research project investigating the

relationship between ASL and English literacy skills (Prinz et al., 1995; Strong & Prinz, 1997, 2000). The TASL allows an in-depth investigation of specific linguistic structures and, thus, does not provide a screening mechanism for deaf children. As of the writing of this review, the TASL has been reported to have been used with 155 deaf students, aged 8–15 years. The TASL consists of two production and four comprehension measures, which are administered individually[2]:

Production measures.

1. Classifier Production Test: A short cartoon movie is shown to the test takers. The cartoon is then presented again in 10 segments. Participants are asked to sign each segment in ASL. The videotapes of their signed responses are scored for the presence of different size, shape, and movement markers in the classifiers.

2. Sign Narrative: Pictures from a children's book (*Good dog Carl*; Day, 1996) without text are given to the participants with the task to tell a story. Their signed versions of the story are videotaped and scored for the use of specific ASL grammar and narrative structures, based on a checklist.

Comprehension measures.

1. Story Comprehension: An ASL narrative presented by a native signer is shown on video. While watching the video, the participants are asked questions about the content. Their responses are videotaped.

2. Classifier Comprehension Test: Pictures of objects with a variety of visual features are shown to the participants. Next, they see a deaf person describing each object in four different ways. Following these descriptions, participants are asked to select among different video still frames in their text booklet the description that best matches the picture stimulus.

3. Time Marker Test: Six representations of a specific time or period of time are shown on video. On a calendar-like answer sheet, the participants are asked to mark the corresponding dates.

4. Map Marker Test: A videotaped description is shown for ways objects are located in different types of environments, for example, vehicles at crossroads or

furniture in a bedroom. For each description, participants are asked to select the correct representation from a selection of photographs in an answer booklet.

Stages of test development.   In the first stage of this project, the TASL was developed, a refinement of data collection procedures was made, sampling procedures were planned, and testing was done on a small sample. The results of this pilot phase indicated that the instrument measurements were both reliable and valid. A draft of the test was sent to five well-known American deaf linguists, who reviewed the test and provided feedback. As a result of this feedback, the test was revised. In the project's second stage, three measurements were conducted with the deaf test participants: the TASL, the *Woodcock–Johnson psychoeducational test battery—revised version* (Woodcock & Mather, 1989), and the *Test of Written Language* (Hammill & Larsen, 1996).

The subjects in this study were 155 deaf students from the same testing site that was used for the pilot study. They were divided into two age groups: 8–11 years old and 12–15 years old. From these 155 deaf students, 40 had deaf parents (in two cases, only one parent was reported deaf) and 115 had hearing parents.

Participants were tested during the school day in two 1-hr sessions. One session was assigned for the TASL and one session for the English literacy test. A deaf researcher fluent in ASL administered the TASL, and test instructions were given in ASL on video. The signed responses were videotaped and later scored by a deaf researcher. Interrater reliability was established for each TASL subtest by having raters score 10 protocols, review them, resolve disagreements, and then score a second set of protocols. Following this approach, the raters reached a high agreement of about 96% (Strong & Prinz, 1997). In order to distinguish between participants' levels of proficiency, the ASL scores were divided into three groups, resulting in low, medium, and high levels.

Psychometrics of the TASL.   For the TASL, the evidence for interrater reliability was reported as 96% (Strong & Prinz, 1997, p. 40) and was thus quite high. As well, content validity was provided, which

includes the feedback by an advisory panel of five well-known deaf linguists on the revised version of the TASL.

Adaptations of the TASL to other sign languages. To this date, the TASL has been adapted for Catalan Sign Language at the Autonomous University of Barcelona in Spain, into Swiss French Sign Language in cooperation with the Bilingual School for the Deaf in Geneva and the Department for Psycholinguistic at University of Geneva, Switzerland (Niederberger, 2004), and into Swedish Sign Language in collaboration with the University of Stockholm (Schönström et al., 2003).

BSL Receptive Skills Test

The BSL project was based on the objective to design, produce, and standardize an assessment instrument for BSL to be used with deaf children (Herman et al., 1999). This type of instrument has long been of special interest to professionals working with deaf children for making baseline assessments, identifying language difficulties, and evaluating the outcomes of therapy programs (Herman, 1998; Herman, Holmes, & Woll, 1998). The BSL Receptive Skills Test is designed for children aged 3–11 years. Following a pilot study on 41 deaf and hearing children between 3 and 11 years (28 with one deaf parent and 13 hearing children with a native signing background), the test was revised and has been standardized on 135 children. The participants in the standardization study included (a) deaf children with deaf parents, (b) hearing children of deaf parents (with a native signing background), and (c) selected deaf children of hearing parents (identified by the teachers) who were enrolled into a bilingual program, had hearing parents with unusual good signing skills, or older deaf siblings. Given this standardization sample, it should be noted that the test norms do not parallel test norms for children who are hearing, where all children are native speakers of the language. Rather, the standardization sample mixes native and nonnative signers, as well as including hearing children who are assumed to be developing BSL in a typical manner.

The BSL Receptive Skills Test focuses on selected aspects of morphology and syntax of BSL. It consists of a vocabulary check and a video-based receptive skills test.

Vocabulary check. The vocabulary measure is designed to ensure that participants understand the signs used in the receptive skills test. The test takers confirm their knowledge of the 22-item vocabulary through a simple picture-naming task that identifies signs, using different pictures borrowed from the receptive skills test.

Video-based receptive skills test. The video-based receptive skills test consists of 40 items, which are ordered by level of difficulty. Due to the regional variation in signs, there are two versions of this task, one for the North and one for the South of the United Kingdom. In this task, deaf participants' receptive knowledge of the BSL structures of syntax and morphology is assessed:

- spatial verb morphology,
- number and distribution,
- negation,
- size/shape specifiers,
- noun–verb distinction, and
- handling classifiers.

The pictures used in this test depict easily recognizable objects and are appealing to children in the targeted age range (3–11 years). Additional distracter items are included to reduce guessing, and the location of the target picture on the page is randomized.

Testing procedures. The receptive skills test is presented to the participant in video format. In addition to the test items, it also includes instructions and test stimuli. This format facilitates a standardized presentation of the test and reduces the demands on the tester. The vocabulary check, however, is administered live and requires some BSL skills on the part of the tester.

Participants are assessed on the vocabulary checklist, the BSL Receptive Skills Test, and other recently published BSL assessments, for example, a BSL Production Test (Herman, et al., 2004).[3] The BSL tests are administered by a deaf researcher with fluent BSL skills.

Psychometrics of the BSL Receptive Skills Test. In order to establish test–retest reliability for the receptive tasks, 10% of the sample based on which the test was

standardized, were retested. The test scores improved on the second testing, but the rank order of scores was preserved. There was also a high correlation (.87) between the test and retest scores. Split-half reliability analysis for the internal consistency of the receptive test revealed a high correlation (.90) and, therefore, represents a high internal consistency. The scores for the BSL Receptive Skills Test of the children involved in the pilot were compared with those of subjects not yet exposed to the test materials. There was a slight advantage in the pilot children; however, the difference between the groups did not achieve statistical significance ($p = .7$).

Adaptations for other sign languages.   The BSL Receptive Skills Test is one of the few instruments to assess deaf students' sign language skills that are commercially available (Herman et al., 1999). So far, it has been adapted for LSF (C. Courtin, personal communication, May 2, 2002), Australian Sign Language (Auslan; Johnston, 2004), Danish Sign Language (DSL; T. Larsen, personal communication, May 9, 2004; January 6, 2005; January 11, 2005), and Italian Sign Language (Lingua Italiana dei Segni [LIS]; Surian & Tedoldi, 2005). An adaptation of this test to German Sign Language has been completed (Haug & Mann, 2005) and the results of the first test data are currently being analyzed.

Reported Linguistic Issues Related to Test Adaptation of the TASL and BSL Receptive Skills Test

The differences between languages fall on a wide continuum that ranges from insignificant to considerable, depending on which languages are being compared. In this respect, comparisons of natural sign languages make no exception. The adaptation of a sign language test from one visual spatial modality into another becomes a complicated task requiring researchers to carefully consider the linguistic differences that exists between both the source and the target sign language (Mason, 2005). What makes this process even more difficult is the overall lack of sign language research in most countries and more specifically, the absence of cross-linguistic research upon which to draw.

To illustrate this point, we will present examples from recent projects for developing tests, all of which involved the adaptation of either the TASL or the BSL Receptive Skills Test.

Linguistic Issues in the Adaptation Process of the TASL

One important linguistic difficulty confronted in the adaptation of a test concerns the categorization of linguistic features. Categorization differences can be of particular importance during the test administration process and might require methodological changes.

Research by Schönström et al. (2003) on the adaptation of TASL for Swedish Sign Language addresses some of the problems the authors encountered while scoring students' performances on the classifier production test based on different categorization of these linguistic features in each sign language. The researchers ended up using a score sheet that had a less detailed breakdown of classifier subcategories and focused on a smaller number of easily recognizable features, instead, to facilitate the scoring process. On this modified sheet, features to be checked included the amount of polymorphemic verbs scored, the number of signs used by a participant, as well as the types of different classifiers.

Linguistic Issues in the Adaptation Process of the BSL Receptive Skills Test

Lexical differences.   Surian and Tedoldi (2005) address the issue of lexical differences in their work on the adaptation of the BSL Receptive Skills Test for LIS. In their adapted version, the total count of vocabulary items came to 21 instead of the 22 signs used in the original BSL test because two of the items (i.e., "boy" and "child") which signers express by using different signs in BSL are represented by the same sign (i.e., CHILD + MALE or FEMALE) in LIS. As a result, the LIS test version only includes a vocabulary card for "boy" which is signed CHILD MALE.

Morphosyntactic issues.   In their adaptation for LIS, Surian and Tedoldi (2005) experienced difficulties related to morphology and syntax, as well, particularly when trying to adapt structures that involved

negation. These difficulties may have stemmed from the wider variety of devices that signers of LIS have at their disposal to express this grammatical feature in comparison to users of BSL.

C. Courtin and his team, while working on an adaptation of the BSL Test for LSF, reported other issues related to negation (C. Cuxac, E. Lawrin, F. Limousin, personal communication, June 10, 2003). In this study, the researchers faced the challenge of working with a smaller number of forms of negation in the target sign language, LSF, than in BSL. Whereas the BSL test consists of 40 items of which eight items represent different forms of negation (e.g., BSL signs such as NOTHING, NO, NOT, NOT-LIKE), LSF has fewer signs to express negation. The effect this discrepancy had on the adopted version for LSF was item redundancy as some items ended up measuring the same forms of negation more than once.

A different morphosyntactic issue was encountered by a team of deaf and hearing teachers in Denmark, working on the adaptation of the BSL Receptive Skills Test for DSL (T. Larson, personal communication, January 6, 2005). This issue occurred in a task related to noun–verb derivation in which the participant had to distinguish between verb and noun forms using the same hand shape but different movements (e.g., like the morphological process of reduplication of the movement which distinguish the ASL signs for CHAIR and SIT). After encountering a number of items (e.g., PENCIL–WRITE) for which DSL uses two completely different signs rather than morphological variation, the Danish team decided to replace the original items with noun–verb pairs that do exist in DSL (e.g., PAINTBRUSH–PAINT) and which follow more closely the morphological processes used in the source test. Despite these modifications to the test, the team continued to question to what extent Danish deaf native signers actually express the difference between noun and verb signs by an alternate type of movement, as found in BSL or ASL.

Research by Johnston (2004) suggests that even in the case of adaptations where both sign languages share the same history (Johnston, 2002), differences may occur. The more recent study, which investigated an adapted version of the BSL Receptive Skills Test (Herman et al., 1999) for Auslan, examined the sign language skills of deaf and hearing students in a bilingual English/Auslan program in Sydney, Australia. Johnston (2004) reports that during the adaptation process from BSL to Auslan, the BSL signs PENCIL–WRITE, which morphologically mark a noun–verb distinction, were substituted by Auslan signs, which represent this noun and verb by unrelated lexical signs.[4]

Cultural Issues Related to Test Adaptation

Language and culture are undoubtedly tightly interconnected across different modalities. Yet, the extent to which these connections have been empirically proven remains limited. Thoutenhoofd (2003) addresses the importance for researchers conducting cross-linguistic studies of taking into account possible differences between the communities using the source and target languages. Some culturally related differences may not be specific to Deaf cultures alone but rather have their roots in cultural aspects of the linguistic majority that are shared by the linguistic minority. Examples of such differences include instructional techniques used in school (Schönström et al., 2003), items that do not exist at all in one culture (Prinz et al., 1995) or only in a modified form (T. Larsen, personal communication). These kinds of cultural differences can present themselves in the size, shape, and color of a British mailbox in contrast to a Danish mailbox and/or in the use of places, objects, or characters that may have varying degrees of familiarity among different cultures. Some of these differences are evident and can be minimized by revising the test materials (e.g., adapting the picture of the mailbox to match the size/color of the country for which the test is used) or even replacing items; others may not be noticed until the analysis of test results has been completed.

Cultural Issues in the Adaptation Process the TASL

Differences in test performance.   "Culture" is a complex concept, thereby making it difficult to find empirical evidence which completely rules out other influential factors on test takers' performances. To what extent can any differences in performances on the

source/target test be related to cultural differences versus linguistic and/or cognitive factors? Moreover, how can such differences in an adapted test version be minimized?

Prinz, Niederberger, Gargani, & Mann (2005) approached these questions in the form of a quantitative–qualitative study to explain the causal impact of culture on sign language ability. They compared selected items from two of the six subtests (i.e., the Time Marker Test and the Story-Comprehension Test) of the original TASL with their adapted versions in for Swiss French Sign Language (Niederberger, 2004) across languages. Results of this measure indicated noticeable differences in participants' responses for one of the items from the story-comprehension task that had to do with obtaining a driver's license. Whereas most American test participants showed no difficulties with this item, it was reported to be one of the harder items for their Swiss French peers. The researchers hypothesized that this divergence may be due to the different significance of having a car in each culture.

Cultural Issues in the Adaptation Process of the BSL Receptive Skills Test

Similar issues related to cultural differences were encountered during the adaptation of the BSL Receptive Skills Test. For example, the Danish team of researchers included in their version pictures that were used in the source test (T. Larsen, personal communication, January 6, 2005). Some of these pictures were difficult for Danish participants to understand and had to be altered in order to match the cultural background of these test takers. In this context, the color (red) and shape (round) of the mailbox that appears on one of the pictures was changed to match the image of a Danish mailbox (yellow, squared) while the location of the steering wheel in a picture depicting a car was changed from right to left side. In another case, the authors decided to replace the picture of a yellow dog with black spots with that of a different dog due to strong similarities of the BSL picture to a character in a well-known Danish children's book. Furthermore, a number of additional pictures were changed because the Danish teachers were not satisfied with the quality of these pictures or felt that they were difficult to com-

prehend (e.g., people going up on an escalator, the action of a boy drinking, and a group of three people representing a queue).

Psychometric Properties in the Adaptation Process

Ensuring good psychometric properties of a sign language test is not only a matter of test development in general but also of test adaptation. As pointed out earlier in this article, the psychometric properties that have been established for one test cannot be adapted or "assumed" to be the same for the target test version. They need to be restandardized. Most of the adapted test versions discussed in this article lack a thorough documentation of psychometric properties. This remains one of the main weaknesses for both sign language test development and adaptation.

Summary and Conclusions

Many problems encountered in the adaptation of assessment tests from one natural sign language to another stem from differences between the test languages or from differences in cultural backgrounds among the test takers. Some of the complications that were based on cultural differences can be solved relatively easy, for example, by modifying stimuli, such as pictures, to fit the target culture. However, such modifications prove to be more difficult when it comes to language-related differences, which may require significant changes in test design.

By providing information on possible issues that may arise when adapting one natural sign language test for another, it was our intention to show that such a process needs to be approached with great caution. The issues that we raise in this article are primarily meant as guiding steps to aide researchers for future studies in this field and prevent them from reencountering obstacles encountered by previous test developers. Clearly, this discussion is far from being complete and it is likely that additional issues may be brought up in future studies.

Because the task of adapting a sign language test is not a straightforward task, it is of particular importance to give the adaptation process ample time and consideration. The assessor needs to carefully weigh the advantages of adaptation (e.g., availability of an

instrument that has already been tested and standardized) against possible shortcomings (e.g., significant language/cultural-related differences between the two sign languages and language users) in order to determine which approach most effectively meets the needs of their test takers' individual situation.

Given the current state of sign language-related research in Europe, especially with regard to assessment, international collaborations of scientific efforts and resources could prove to be useful. Such joint efforts may help to reduce some of the linguistic as well as culturally based issues described in this review, due to the close cultural ties between many European countries. In addition, a cooperative approach could facilitate any efforts to develop future sign language assessment tools based on a template which incorporates the individual features of each natural sign language.

Notes

1. For a comprehensive review of several sign language tests, visit the following Web site: http://www.signlang-assessment.info

2. Currently, TASL is being revised to become a Web-based diagnostic tool to be used by schools. For this article, the authors refer to the original TASL version from 1995.

3. For more information on these other assessments, see http://www.city.ac.uk/lcs/compass/bsldevelopment/assessingbsldevelopment.html

4. Similar issues were addressed in a study by Schembri et al. (2002) during the adaptation from Auslan of the Test for ASL Morphology and Syntax (Supalla, Newport, Singleton, Supalla, Coulter, & Metlay, 1995).

References

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). Dictionary of language testing. In M. Milanovic (Series Ed.), *Studies in language testing* (Vol. 7). Cambridge, UK: Cambridge University Press.

Day, A. (1996). *Good dog Carl*. Simon & Schuster Inc.: New York, NY.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test—Third edition*. Minneapolis, MN: Pearson Assessments.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6,* 304–312.

Hammill, D. D., & Larsen, S. C. (1996). *TOWL-3—Test of written language: Third edition*. Austin, TX: PRO-ED.

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 [Surveying the need for sign language tests for hearing-impaired children]. *Das Zeichen, 64,* 220–229.

Haug, T., & Mann, W. (2005). Projekt Gebärdensprachtest—Entwicklung von Testverfahren zur Deutschen Gebärdensprache für gehörlose Kinder und Jugendliche [Project sign language test—Development of test instruments for German sign language for deaf children and adolescent]. *Das Zeichen, 71,* 370–380.

Herman, R. (1998). The need for an assessment of deaf children's signing skills. Deafness and education. *Journal of the British Association of the Teachers of the Deaf, 22,* 3–8.

Herman, R., Grove, N., Holmes, S., Morgan, G., Sutherland, H., & Woll, B. (2004). *Assessing BSL development: Production test (narrative skills)*. London: 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*. Unpublished manuscript, City University London, UK.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL development—Receptive skills test*. Coleford, UK: The Forest Bookshop.

Johnston, T. (2002). BSL, Auslan and NZSL: Three signed languages or one? In A. Baker, B. van den Bogaerde, & O. Crasborn (Eds.), *Proceedings of the Seventh International Conference on Theoretical Issues in Sign Language Research* (pp. 47–69). Hamburg, Germany: Signum Verlag.

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

Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.

Mann, W., & Prinz, P. (2006). The perception of sign language assessment by professionals in deaf education. *American Annals of the Deaf, 151,* 356–370.

Mason, T. C. (2005). Cross-cultural instrument translation: Assessment, translation, and statistical applications. *American Annals of the Deaf, 150,* 67–72.

Mayberry, R., & Eichen, I. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language, 30,* 486–512.

Mayberry, R., Lock, E., & Kazmi, H. (2002). Development: Linguistic ability and early language exposure. *Nature, 417,* 38.

Morford, J., & Mayberry, R. (2000). A reexamination of "early exposure" and its implications for language acquisition by eye. In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 111–128). Mahwah, NJ: Lawrence Erlbaum Publishers.

Niederberger, N. (2004). Capacités langagières en Langue des Signes Française et en français écrit chez l'enfant sourd

bilingue: quelles relations? [Linguistic proficiency of the deaf bilingual child in French Sign Language and written French: is there are a relationship?]. Simon & Schuster Inc.: New York, NY.

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

Prinz, M., Strong, M., & Kuntze, L. (1995). *A test of ASL.* Unpublished manuscript, San Francisco State University, California Research Institute.

Prinz, P., Niederberger, N., Gargani, J., & Mann, W. (2005, July). *Cross-linguistic and cross-cultural issues in the development of tests of sign language.* Paper presented at the International Congress for the Study of Child Language (ICSL), Berlin, Germany.

Rapp, J., & Allalouf, A. (2003). Evaluating cross-lingual equating. *International Journal of Testing, 3,* 101–117.

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.

Schönström, K., Simper-Allen, P., & Svartholm, K. (2003, May). Assessment of signing skills in school-aged deaf students in Sweden. In *European days of deaf education* (pp. 88–95). Örebro, Sweden.

Singleton, J. L., & Supalla, S. (2003). Assessing children's proficiency in natural signed languages. In M. Marschark & P. Spencer (Eds.), *Oxford handbook of deaf studies, language and education* (pp. 289–302). New York: Oxford University Press.

Solano-Flores, G., Trumbull, E., & Nelson-Barber, S. (2002). Concurrent development of dual language assessments: An alternative to translating tests for linguistic minorities. *International Journal of Testing, 2,* 107–129.

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.

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

Supalla, T., Newport, E., Singleton, J., Supalla, S., Coulter, G., & Metlay, D. (1995, April 20). *An Overview of the test battery for American Sign Language morphology and syntax.* Paper presented at the Annual Meeting of the American Educational Research Association (AERA). San Francisco, CA.

Surian, L., & Tedoldi, M. (2005). *Adaptation of BSL receptive skills test to Italian sign language (LIS).* Unpublished manuscript, University of Trieste, Italy.

Thoutenhoofd, E. D. (2003). Inclusion of deaf pupils in standardised educational assessments: Potential sources of differential item functioning (DIF). *Deaf Worlds, 19,* 49–78.

Woodcock, R. W., & Mather, N. (1989). *Woodcock-Johnson psycho-educational test battery—Revised.* Allen, TX: DLM Teaching Resources.