Building an Assessment Use Argument for sign language: the BSL Nonsense Sign Repetition Test

Wolfgang Mann & Chloe Marshall

Department of Language and Communication Science, City University London

and the ESRC Deafness, Cognition and Language Research Centre, University College London

Word Count: 5441 (main text)

Corresponding author:

Dr. Wolfgang Mann

Department of Language and Communication Science

City University

Northampton Square

London EC1V 0HB

Email: wolfgang.mann.1@city.ac.uk

Abstract:

In this article, we adapt a structural concept introduced for validation in educational measurement, the *Assessment Use Argument* (Bachman, 2005), as a framework for the development and/or use of sign language assessments for Deaf children who are taught in a sign bilingual education setting. By drawing on data from a recent investigation of Deaf children's nonsense sign repetition skills in British Sign Language (Mann, Marshall, Mason, & Morgan, forthcoming), we demonstrate the steps of implementing the *Assessment Use Argument* in practical test design, development and use. This approach provides us with a framework which clearly states the competing values and which stakeholders hold these values. As such, it offers a useful foundation for test designers, as well as for practitioners in Sign Bilingual Education, for the interpretation of test scores and the consequences of their use.

Keywords: assessment, sign bilingualism, phonology, non-word repetition, Deaf Education

Introduction: Sign Bilingual Education in the UK

Over the course of the past decade, Sign Bilingual Education in the UK has become more established, due to changes in the educational context and to an increased understanding of sign language development based on research (Swanwick & Gregory, 2007). While the number of Deaf children attending sign bilingual education programs[1] which use British Sign Language (BSL) as (primary) means of instruction and communication is relatively small, this does not necessarily represent the true number of children who are sign bilingual. In fact, many Deaf children who attend mainstream programs or oral schools with little or no use of sign language are likely to be exposed to sign language outside the classroom (or may themselves be native signers), through interaction with other Deaf peers or at home when communicating with Deaf family members. Efforts to gather reliable information are impeded by the lack of available census information on Deaf children's type of schooling. Additional challenges lie in the heterogeneous nature of this target group: there is significant variation across individuals with regard to the degree of hearing loss, age of language onset, communication at home, linguistic and cultural background, etc. (Andersen, 2006; Humphries & Allen, 2008; Marschark, Lang, & Albertini, 2002). As a result, practitioners' decisions on issues including placement and type of intervention need to be made on a case-by-case basis rather than following a 'one approach fits all' procedure.

[1] Throughout this paper, we are referring to Swanwick & Gregory (2007), who define sign bilingualism as an "approach to the education of Deaf children which includes sign language as means of instruction and communication" (9).

In light of these demands, credible evidence is indispensable for supporting decision-making, and this requires appropriate assessments. For instance, in a sign bilingual education environment, special consideration needs to be given to the question how Deaf pupils' sign language can be properly assessed. In addition to designing tests that measure different aspects of an individual's sign language proficiency, more explicit links are needed between test scores, their interpretation, and the consequences of their use (e.g., type of therapy the child receives, type of educational placement, etc.). These links are essential for maximizing the efficiency (not only with regard to quality but also whether the test will be consistently used long term) of the assessment and for meeting the requirements of the key stakeholders (e.g., practitioners, Deaf students, school administrators). The lack of such links creates a gap between test design and test use, leaving each area more susceptible to errors and/or misguided decisions.

Given the relative youth of sign language research and the limited number of studies that have focused on sign language assessment, this area needs to be approached with extra caution, if some of the mistakes from regular education (e.g., treating test validity and test use as unrelated issues) are to be avoided. Because many Deaf children receive only inconsistent or incorrect language input (Kuntze, 1998) and some may not have exposure to any language until they enter school (Moores, 2001), they are at a constant disadvantage compared to their hearing peers (Kuntze 1998; Meier & Newport 1990). Consequently, there is a need for appropriate assessment based on which decisions regarding suitable intervention can be made. In order to avoid further delays in students' language development, test-developers and test-users[2] need to collaborate on constructing a framework for the assessment. This includes discussing the primary aims of the assessment, including practicability, reliability, level of interactivity, and validity. Because test-developers and test-users often approach collaborations with different expectations about the purpose of assessment, agreeing on these aims is an important prerequisite for successful links between inference and consequence of testing. The main idea behind the close collaboration between test-developers and test-users is to set up a framework for quality control during the development phase and continue to examining the overall usefulness throughout the test cycle. Following the *Assessment Use Argument (AUA)* approach, conceptualized by Bachman (2005), helps to further specify links between test validity and test use.

In this paper, we present some of the challenges of sign language assessment related to assessing Deaf children. We then introduce the *AUA* and provide some background information on the data from a recent study on investigating Deaf children's phonological skills which adapts a non-word repetition methodology for sign language. Finally, we explore the suitability of the *AUA* for sign language assessment by drawing from the aforementioned data on our Nonsense Sign Repetition Task.

2. Some of the challenges of sign language assessment

Language assessments are frequently used by professionals across disciplines, including education, psychology, and linguistics (Bachman & Palmer, 1996), for a

[2] In the context of this paper, we make a distinction between test-users and test-takers. The first group includes test administrators and practitioners, the second group includes test participants.

number of reasons. These reasons include documenting students' developmental progress in school, measuring performance of second language learners, diagnosing error patterns in late language learners, and/or serving as a linguistic research tool. In the field of education, language assessments often play a key role not only in measuring students' academic success but also to rank schools. Generally, these assessments have been developed for, and normed on, hearing pupils and are not appropriate for the particular needs of most Deaf test-takers.

Compared to spoken languages, research on the assessment of sign language can be considered still in its infancy and, while several assessments have been developed, most of these instruments have not been standardized or are used mainly for linguistic research (e.g., *Test Battery for American Sign Language Morphology and Syntax*, Maller, Singleton, Supalla, & Wix, 1999; *Test Battery for Australian Sign Language Morphology and Syntax*, Schembri, Wigglesworth, Johnston, Leigh, Adam, & Baker, 2002; *Test for Grammatical Judgment of ASL*, Bordreault & Mayberry, 2006). Notable exceptions include the *BSL Receptive Skills Test* (Herman et al, 1999) and the *BSL Productive Skills Test* (Herman et al, 2004). For many practitioners in Sign Bilingual Education, this means they either have to rely on standardized assessments developed for hearing pupils[3] or make up their own informal assessments, if they want to measure Deaf pupils' language proficiency. Some of the most relevant challenges of sign language assessment within the

[3] For example, to generate lists for assessment of vocabulary in BSL, practitioners may adapt selected items from well-established tests, including the *Peabody Picture Vocabulary Test (PPVT4FA)*(Dunn & Dunn, 2007) or the *Expressive One Word Picture Vocabulary Test (EOWPVT-2000)*(Brownell, 2000).

context of Sign Bilingual Education are discussed next (for a more general review of sign language assessment, the reader is referred to Haug & Mann, 2008).

The first and most apparent challenge is the overall scarcity of available, standardized sign language tests, specifically those that are appropriate for use in an educational setting (e.g., Haug & Mann, 2008; Hoffmeister & Schick (forthcoming); Jamieson, 2003; Mann, 2008; Prezbindowski & Lederberg, 2003). In this context, relevant aspects include test length, time for scoring, and the linguistic knowledge required of the scorer (Singleton & Supalla, 2003). An example is the *Test Battery for American Sign Language Morphology and Syntax* (Maller et al, 1999), which requires administration and scoring by a Deaf native signer and takes 2 hours to administer and 15 hours to score (Maller et al, 1999).

Another challenge lies in the limited time and resources available for sign language test development, specifically for practical use. Similar to spoken languages, these constraints are often tied to the lack of clear mechanisms for integrating existing lists of "more or less independent qualities and questions into a set of procedures for test-developers and users to follow" (Bachman, 2005, 1).

Finally, one of the key challenges to successful sign language assessment, which applies equally to spoken languages, is the lack of an overall framework to link assessment performance to use (i.e., decision-making) (Bachman, 2005). This requires a transparent connection between test performance and interpretations, and from interpretation to use. One of the possible advantages of such a connectional framework would be to apply data collected mainly for research purposes, where appropriate, to explore possible uses for practitioners in Sign Bilingual Education

In this paper, we will focus in more detail on the last key challenge. The need for sign language assessment tests, in particular for use with young test-takers, has been well established in the literature (e.g., Haug & Hintermair, 2003; Herman, 1998; Mann & Prinz, 2006). In a series of recent studies, researchers have developed a number of language assessments specifically for use with Deaf children. The areas investigated in these studies included Deaf children's phonological skills (Mann, Marshall, Mason & Morgan, forthcoming), their speech-reading ability (Kyle et al, 2007), and their sentence processing skills (Mason & Rowley, personal communication). Given the emphasis of these studies on assessment, they make ideal examples to explore the suitability of the *Assessment Use Argument*. For the purpose of this paper, we will draw on data from one study in particular (Mann et al., forthcoming), which investigated Deaf children's phonological skills in BSL.

3. The Assessment Use Argument

The structural concept for sign language assessment that we refer to in this paper is the *Assessment Use Argument* (*AUA*). The *AUA* is grounded in research focusing on test validity, which has explored the interaction between different areas of language ability that a test-taker draws on during (spoken) language assessment. It moves away from the traditional approach to language assessment, i.e., providing sets of procedures for investigating and supporting claims about score-based inferences without addressing issues of test use or consequences of test use. Instead, the *AUA* suggests approaching assessment by first setting up a framework in which the different processes associated with assessment are made visible and are connected. These processes represent two major levels of assessment, i.e., validity and utilization; they include measurement of performance, interpretation of performance scores, as well as decisions related to performance. Both utilization and validity form sub-arguments. They are supported by claims, or interpretations, we want to make "on the basis of the data, about what a test-taker knows or can do" (Bachman, 2005, 9). Fig. 1 illustrates how the validity argument and the utilization argument can be connected to form the *AUA*, using Toulmin's (2003) argument structure (in which arguments consist of claims, supported by data and warrants).

//Insert Figure 1 about here//

At each of the two argument levels, the structure includes data and general statements used to provide legitimacy (warrants) to generate the claims. Furthermore, each warrant is supported by other assurances (backing), which may include theories, findings from previous research, or other types of evidence gathered as part of the test validation process. During the process of gathering data in support of an argument, it is possible that alternative explanations or counterclaims (rebuttal) to the intended interference are encountered, which may be supported, weakened, or rejected based on the data (Bachman, 2005). For the assessment utilization argument, the number and/or type of warrants are flexible and may vary from argument to argument while for the assessment validity argument, there is only one warrant.

The aim behind Bachman's *AUA* is to provide a framework to guide and facilitate the development, use and evaluation of assessment. This is done by combining the two sub-arguments, utilization and validity, into one argument, and connecting the different functions and uses of assessment (e.g., data interpretation, consequences of assessment, assessment use) to maximize the effectiveness of the assessment instrument. This combination generates some kind of assessment protocol to guide both test-developers and test-users.

Our aim in this paper is to explore the suitability of the *AUA* for Sign Bilingual Education. We do so by using data from a recent study in which we investigated Deaf children's performance on a Nonsense Sign Repetition Test. A brief description of the background and methodology of this test follows next.

Background

To this point, there has been little research on signers' acquisition of phonological[4] skills, and the majority of existing studies are single or small-scale case studies (e.g. Morgan, 2006; Meier, Mauk, Cheek & Moreland, 2008). There have been few descriptions of sign phonology development in older Deaf children, and none of the existing research has compared Deaf children's general sign language skills with their developing phonological abilities. Despite this, several studies have shown similarities in the development of sign phonology to previously documented cases in the speech literature (Boyes-Braem, 1990; Clibbens & Harris, 1993; Meier, 2005; Morgan, 2006). In particular, structural complexity affects phonological acquisition in signed languages as in spoken languages, with young children simplifying phonological forms and mastering complex target forms only gradually.

In our study (Mann et al, forthcoming), we investigated Deaf children's phonological skills in BSL, specifically their ability to repeat nonsense signs. This

[4] "Phonology" is a sublexical level of structure consisting of patterns of meaningless units (sounds in spoken languages, and gestural units in signed languages). Phonology is in part constrained by the physiology of the systems involved in production and perception (the oral-auditory channel in spoken languages, and the visual-gestural channel in signed languages).

research was motivated by a non-word repetition methodology which has been widely used in spoken language research (for reviews, please see Coady and Evans, 2008, and Gathercole, 2006). In these tasks, the participant listens to a set of nonsense words (i.e. words that are phonologically possible but do not have any meaning) and repeats each one immediately after hearing it. Skills that are measured include both perception and production. In addition, the task assesses the ability to encode a phonological representation for storage in phonological working memory and to retrieve it from there (Gathercole, 2006). Because the child has never heard the items before, the task taps into the child's productive phonology, unconfounded by stored lexical knowledge. Non-word repetition abilities have been linked to word-learning abilities (Gathercole, 2006) and to language development more generally. For instance, children with Specific Language Impairment (SLI) and dyslexia have difficulty repeating non-words (Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990), and non-word repetition ability is an excellent predictor of language learning ability in children learning English as a second language (Service, 1992).

This methodology was used for the present study for two reasons: firstly, to allow us to systematically investigate the development of phonological skills in a large group of Deaf children, and secondly to create a test of phonological working memory that is suited to sign language. We discuss each of these points in turn.

Sign Language Phonology

Comparable to spoken languages, signed languages systematically organize meaningless phonological units into meaningful ones (Stokoe, 1960; Brentari, 1998) Signs are made up of three basic phonological categories or “parameters”: handshape, movement and location. These terms are fairly self-explanatory. ‘‘Handshape’’ denotes the particular shape that a hand makes in a sign, and handshapes vary in the number of fingers that are selected and how these fingers are flexed or extended. Approximately 50 different handshapes are attested in BSL (Brien, 1992). There are two classes of movement - ‘‘path movements’’, which involve movement of the hand and arm, and ‘‘hand-internal movements’’ that involve just the fingers or wrist. Signs contain either one or both of these types of movement. An example of a BSL sign containing two movements is FIRE (noun), which consists of an up-and-down movement of both hands while, at the same time, the fingers move back and forth (wiggle). Most signs are produced in a neutral location in front of the signer, but they can be produced elsewhere, such as at various locations on the face and torso.

Handshapes and movements differ in their complexity, and therefore in how easily they are acquired. Simple handshapes that are easy to articulate, such as a fist (“A”[5]), a fist with the index finger extended (“G”), or an open hand (“5”), are amongst the first to be acquired. More complex handshapes, such as “Y”, with the thumb and little finger extended, and “W”, where the thumb and little finger are bent and the other three fingers extended, are acquired later, and young children will often use simpler handshapes in their place (Morgan, 2006; Morgan, Barrett-Jones & Stoneham, 2007). With respect to movement, signs with both a path and internal movement, termed a “movement cluster”, cause difficulties in acquisition, with children sometimes deleting one of the movements, or producing the two sequentially rather than simultaneously

[5] We follow the convention of naming handshapes after the letters they represent in the American Sign Language manual alphabet or the numbers they represent in the counting system

(Morgan, 2006; Morgan, Barrett-Jones & Stoneham, 2007). By contrast, location represents by far the simplest part of the sign for Deaf children to acquire, with very few errors reported after 3 years of age (Cheek, Cormier, Repp & Meier, 2001; Meier, 2005).

Phonological working memory

Phonological working memory (PWM) is the type of short-term memory we use whilst producing and understanding language, be that language spoken or signed. A central feature of PWM is that it is limited, allowing for only a small number of linguistic items to be temporarily maintained and manipulated, and it therefore creates a bottleneck for language processing (Lewis, Vasisth & van Dyke, 2006). Children's PWM capacity increases with age and can be measured in two ways: using span tasks, whereby the participant has to repeat sequences of words or digits forwards or backwards, or non-word repetition tasks.

Digit span tasks have become the focus of considerable debate in the recent sign language research literature. While the average forwards digit span for hearing adults using a spoken language is 7 +/-2, for Deaf and hearing adults using a signed language this is significantly lower, at around 5 +/-1 (Boutla, Supalla, Newport & Bavelier, 2004; Bavelier, Newport, Hall, Supalla & Boutla, 2006). However, although signers are unable to remember as many items in sequence as speakers are, they *are* able to recall the same number of items when their exact sequence is not required (Bavelier, Newport, Hall, Supalla & Boutla, 2008). Hence it appears that sequence is important for PWM in spoken languages, but that sequentiality does not play as great a role in signers' PWM (Bavelier et al, 2008; Geraci, Gozzi, Papagno & Cecchetto, 2008). This conclusion therefore raises the possibility that digit span is not a fair measure of PWM in signers, and that a nonsense sign repetition test, because it takes into account the greater degree of simultaneity in sign language structure, would have more validity. Furthermore, in spoken languages the relationship between non-word repetition and language is stronger than that between digit span and language (Gathercole et al, 1994), thereby reinforcing the validity of non-word repetition as a measure of PWM.

Design

Our Nonsense Sign Repetition Task consisted of 40 nonsense signs all of which were phonotactically possible but meaningless in BSL. Because signs generally contain only one syllable, stimuli could not be created by manipulating the number of syllables, as is most often the case for non-word repetition tests (e.g. the *Children's Test of Non-word Repetition*, Gathercole et al, 1994; and the *Non-word Repetition Test*, Dollaghan & Campbell, 1998). Instead, we manipulated the complexity of the handshape and movement in a 2 x 2 design, as shown in Table 1:

//Insert Table 1 about here//

Items contained handshapes that were either simple or complex. Simple handshapes were 'B', '5', 'G' and 'A' (Sutton-Spence & Woll, 1999). All other handshapes, which were selected from the BSL inventory, were classed as 'complex'. One movement – either internal movement or path movement – was classed as 'simple' and two movements (internal movement plus path movement) as 'complex' while balancing across conditions for different types of path movement (e.g., straight, arc) and different types of internal movement (e.g., opening, closing, wiggling). We also controlled for phonological properties that were not experimentally manipulated, such as one-handedness versus two-handedness.

The design for the Nonsense Sign Repetition Test is based on the theoretical concept that the greater the degree of phonological complexity within a nonsense sign, the greater the load on phonological working memory, and therefore the more difficult the sign will be for the test-taker to repeat (see Fig 2).

//Insert Fig. 2 about here//

The stimuli were produced by a Deaf fluent signer, sitting against a blue screen facing a digital camera. All items were presented to participants as 10 x 14 inch images on a laptop computer with a 15 inch screen.

Participants

91 congenitally Deaf children (60 boys/31 girls) participated in the experiment. They were divided into three age groups: 3-5 years old (N = 26, mean = 4;11, range = 3;4-5;11), 6-8 years old (N = 26 mean = 7;4, range = 6;0-8;10) and 9-11 years old (N = 38 mean = 10;3, range = 9;0-11;9). Participants were recruited through schools for the Deaf in the UK. They were either born into BSL-using Deaf families (N=14) or had very early exposure to BSL at nursery school, and subsequent typical language development as measured using the *BSL Receptive Skills Test* (Herman et al, 1999).[6] Pupils with additional special educational needs and children whose non-verbal cognitive development was below normal were not included in the study.

[6] This test assesses the comprehension selected aspects of BSL morphology and syntax (e.g. negation, number and distribution, verb morphology and the distinction between nouns and verbs) in a picture-pointing paradigm.

Procedure

Each participant was tested individually by a fluent BSL signer in a quiet room at the school in a single session which took between 15-20 minutes. We used pre-recorded instructions by a Deaf native-signing adult, who explained to participants that they were to be presented with a number of novel signs and had to copy each sign as accurately as possible. These instructions were followed by three practice items during which participants could ask questions, if necessary. Once the test began, no more questions were answered. Each stimulus item was presented just once, and the order in which the items appeared was randomized across participants.

Scoring

Responses were scored as either correct or incorrect. Errors were classified according to whether they appeared in the parameters that we had experimentally manipulated (i.e. hand shape, path movement and internal movement), based on a coding scheme developed during the piloting of this test (Marshall, Denmark and Morgan, 2006).

All scores were coded separately by two hearing experimenters, both of whom are fluent signers. These scores were compared and any discrepancies resolved. As additional measure for inter-rater agreement, fifteen participants were randomly selected and coded by a third coder, who was a Deaf native signer. Inter-rater agreement was high (85% for overall score, 88% for hand shape, 87% for path movement, and 83% for internal movement).

Summary of results

Results are reported in full in Mann et al (forthcoming); we present a summary of the most important findings here. Children's accuracy at repeating nonsense signs improved with age, showing that the task captures developmental increases in phonological working memory capacity. Although the number of errors decreased with age, children of all ages made more errors on handshape and internal movement than on path movement, and this replicates previous results from case studies (Meier, 2005; Morgan, 2006). Children were least accurate at repeating the phonologically most complex nonsense signs, although there was no significant difference between their accuracy on the most simple signs and signs with just one level of complexity.

As well as testing children on the Nonsense Sign Repetition Task, we also asked them to complete a bead threading task. This requires children to thread 15 large beads onto a string as quickly as they can, and provides a measure of fine motor control and eye-hand co-ordination. We found a link between nonsense sign repetition accuracy and speed of bead threading for Deaf children under the age of six, even when age was partialed out of the correlation, indicating that for young children fine motor skills are related to sign language production. In addition, we had *BSL Receptive Skills Test* scores available for 65 of the 91 children. Scores on this test correlated significantly with nonsense sign repetition accuracy even age was partialed out, indicating that the Nonsense Sign Repetition Task taps into general BSL language skills.

In order for teachers to make use of the assessment, we established norms for each age group. This was done by subdividing the wide age range (3-11 years) into six groups. Despite the fairly small numbers in each group, it was considered important to maintain yearly age intervals for the younger children (3 years to 5;11) as progress in language development in this period is particularly marked. For the older children (6 years to 11;11 years), two-year age groups were selected. This allowed for larger subject numbers in each group, producing a more reliable basis for the standardization. Participants' raw scores were converted to standard scores and a language quotient was selected as being an easily accessible method of displaying standard scores, using a mean of 100 and standard deviation of 15. These norms have subsequently been made available to the participating schools as part of the Nonsense Sign Repetition Test package.

4. Linking assessment performance to use

At this point, we want to explore the suitability of the *AUA* for sign language assessment, by drawing from the data on the Nonsense Sign Repetition Task (see Fig. 3).

//Insert Fig. 3 about here//

One of the main advantages of the *AUA* is that it can be approached in more than one way, depending on whether the user looks from the perspective of a test developer (bottom to top) or from the perspective of the test user (top to bottom). Because the Nonsense Sign Repetition Task was originally developed as a research tool, we are taking a 'bottom to top' approach in the following discussion.

Starting at the bottom of the figure, we can see the first part of the *AUA*, which is concerned with gathering evidence for the validity of the assessment by linking assessment performance to an interpretation. Within the context of our data, this means that the interpretation of Deaf test-taker's performance on the Nonsense Sign Repetition Task is a valid measure of his or her phonological working memory (PWM). This is backed up by the data from a large number of research studies of non-word repetition in spoken languages (see Coady & Evans, 2008).

A potential rebuttal of this warrant is that the test does not actually tap PWM. BSL is a visuo-spatial language: perhaps Deaf children use visuo-spatial working memory rather than PWM in this task. In order to test this rebuttal, we recruited 46 hearing children aged between 6 and 11 years of age, who had no previous exposure to any sign language, to perform the Nonsense Sign Repetition Task (see Mann et al, forthcoming, for further details). The only way that hearing children could tackle the task was by using visuo-spatial working memory to process what, to them, were non-linguistic gestures. Upon comparison, our results showed that, overall, hearing children performed significantly lower than Deaf children, suggesting that Deaf children were approaching the task differently, and processing the stimuli linguistically using PWM. Based on this evidence, we conclude that the test does indeed tap PWM.

Moving up the figure, we see that four statements (warrants) form the part of the *AUA* which is concerned with utilization, i.e., the link between score-based interpretations and the decisions based on these interpretations. In the case of the Nonsense Sign Repetition Test, the claim is practitioners' decision as to whether it is necessary to provide special intervention to promote development of phonological skills in students with low phonological working memory.

The first statement is concerned with the relevance of the assessment, i.e., the extent to which the ability assessed is a requisite part of the competence. With regard to the Nonsense Sign Repetition Test, our findings comparing Deaf and hearing children show that the ability to successfully repeat nonsense signs does rely on PWM. Furthermore, as discussed in the introductory section on PWM, digit span tasks, which are also used to measure PWM in spoken languages, are not well suited to assessing PWM in sign languages because of their reliance on temporal order. Nonsense sign repetition is more relevant to measuring PWM in a modality whereby language is processed with a greater degree of simultaneity than is the case for spoken languages. Note that this first statement and its backing also provide support for the validity argument, thus reinforcing the link between use and validity (Bachman, 2005).

The second statement addresses the effectiveness of the assessment, or, the extent to which the score-based interpretation can provide information to be used by practitioners and administrators to make appropriate decisions. In case of the Nonsense Sign Repetition Test, the warrant is that repetition accuracy is correlated with wider BSL abilities. This is backed up by our finding of a significant correlation between Deaf participants' repetition accuracy and their score on the *BSL Receptive Skills Test*. This relationship is statistically significant even when age is accounted for in the analysis, meaning that it is not the case that the relationship is driven by older children being better at both tests. Additional support comes from research with hearing individuals which shows that test-takers with high scores on the non-word repetition task have better overall language skills than those who achieve lower scores.

The next statement is related to quality, describing the beneficial outcomes that the test-user expects to achieve by using the assessment[7]. With regard to sign language assessment, the Nonsense Sign Repetition Test enables practitioners to use the results to inform/guide their decisions about Deaf children's language proficiency (in accordance/agreement with other assessments) and determine appropriate intervention measures (where necessary). In addition, Deaf children, who take the test potentially benefit from receiving better support services at school, in particular when scores are low and there is a suspicion of delayed or impaired language, for example Specific Language Impairment (SLI). The statement is supported by previous research with hearing individuals showing that low non-word repetition scores are a reliable indicator of SLI, which, in response, enables practitioners to initiate special intervention services (see review in Coady and Evans, 2008). A study is currently under way to investigate if the Nonsense Sign Repetition Test can be used to identify Deaf children with SLI (Mason & Rowley, personal communication).

The final statement concerns sufficiency of the information provided by the assessment to make a decision. This can be related on what gets included in the definition of the construct, or, ability to be assessed. In the case of the Nonsense Sign Repetition Test, the underlying construct taps into the child's productive phonology, unconfounded by stored lexical knowledge, by measuring perceptual and production skills, as well as the ability to encode a phonological representation for storage in phonological working

[7] Bachman points out that there may always be the possibility, at least in theory, that adding one more assessment will provide more complete information about the ability of interest (2006, 19). This is even more so the case with regard to the assessment of sign language skills in Deaf children about which we still have a fairly limited understanding.

memory and to retrieve it from there. However, the construct by definition does not include lexical knowledge because the signs are not real lexical items, nor does the construct measure morphological or syntactic skills (although it is significantly correlated with them).

Together, the four statements provide support to the decision to be made by practitioners. Still, there are a number of possible counterclaims (rebuttals) to these statements. One potential rebuttal concerns the wide range of scores in any particular age band. This means that a child has to achieve a very low score in order to fall outside the normal range and this may therefore reduce the sensitivity of the assessment in identifying children with real impairments in phonology and phonological working memory. Part of this variability in scores may be linked to higher performance in Deaf children with Deaf parents compared to Deaf children of hearing parents. The former population is much smaller: in our study we had only 14 participants with Deaf parents and therefore not enough to analyze across the wide range of ages that we tested.

Another potential rebuttal concerns the scoring of the test. In its current form, the test is scored in two ways: Each repetition is scored as being correct (awarded 1 point) or incorrect (0 points). The scorer can then fill in a more detailed score sheet indicating exactly where any error or errors have occurred – for example, whether a movement has been deleted, or a handshape substituted. This in-depth error-scoring is time-consuming (between 1- 2 hours) and requires good phonological knowledge of BSL on the part of the scorer, but it has the advantage of providing extra detail as to the child's phonological abilities. The simplified version of the scoring sheet, while quicker (less than 1 hour) and easier to fill out, may not provide sufficient enough information to inform practitioners' decisions. To further investigate this, a number of selected schools will be provided with a trial version of the Nonsense Sign Repetition Test which contains a simplified version of the shore sheet and asked to provide feedback on ease of use as well as sufficiency of results for decision making. On the positive side, the test itself is quick to administer, and even with detailed scoring the total administration plus scoring time takes no longer than 2 hours, considerably less than the time required for administration and scoring of many other sign language assessments.

5. Conclusion

In this paper, we adapted a structural concept for test validation as framework for the development and/or use of a test to assess Deaf children's nonsense sign repetition skills in British Sign Language (BSL). The *AUA* shows potential for use in a sign bilingual (education) context in that it can inform practitioners on decisions regarding the type of intervention most suitable for Deaf students with limited phonological skills in BSL while, at the same time, offering a transparent framework for researchers developing sign language assessments. By fostering a close collaboration between test-developers and test-users, the *AUA* offers an approach to sign language assessment which may help to more efficiently detect and address some of the challenges Deaf children experience as a result of delayed access to language.

Acknowledgements

This work was supported by the Economic and Social Research Council of Great Britain (Grant RES-620-28-6001), Deafness, Cognition and Language Research Centre (DCAL), a City University Research Fellowship awarded to Wolfgang Mann and a Leverhulme Early Career Fellowship awarded to Chloe Marshall. The authors would like to thank all the children and teachers who took part and supported the research.

References

Andersen, D. (2006). Lexical development of Deaf children acquiring signed languages. In B. Schick, M. Marschark, and P.E. Spencer (Eds.), *Advances in the sign language development of Deaf children* (pp. 135-161), Oxford University Press.

Bachman, L. (2005).Building and supporting a case for test use. *Language Assessment Quaterly*, 2(1), 1-34.

Bachman, L. & Palmer, A.S. (1996). *Language Testing in Practice.* Oxford University Press. 1996

Bavelier, D., Newport, E., Hall, M., Supalla, T. & Boutla, M. (2006). Persistent difference in short-term memory span between sign and speech. *Psychological Science*, 17, 1090-1092.

Bavelier, D., Newport, E., Hall, M., Supalla, T. & Boutla, M. (2008). Ordered short-term memory differs in signers and speakers: Implications for models of short-term memory. *Cognition*, 107, 433-459.

Boudreault, P. & Mayberry, R. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, 21(5), 608-635.

Boutla, M., Supalla, T., Newport, E. & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience*, 7, 997-1002.

Boyes Braem, P. (1990). Acquisition of the handshape in American Sign Language: A preliminary analysis. In V. Volterra & C. J. Erting (Eds.). *From gesture to language in hearing and Deaf children* (pp. 107–127). Washington, DC: Gallaudet University Press.

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

Brien, D. (Ed.) (1992). *Dictionary of British Sign Language*. London, UK: Faber and Faber.

Brownell, R. (Ed.) (2000). *Expressive One-Word Picture Vocabulary Test (EOWPVT–2000)*.Academic Therapy Publications.

Cheek, A., Cormier, K., Repp, A., & Meier, R.P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. *Language*, 77, 292–323.

Clibbens, J. & Harris, M. (1993). The acquisition of formational parameters in British Sign Language: a case study. In D. Messer & G. Turner (Eds.), *Critical aspects of language acquisition and development* (pp 197 - 208). London: Macmillan.

Coady, J. & Evans, J. (2008). Uses and interpetations of non-word repetition tasks in children with and without SLI. *International Journal of Language and Communication Disorders*, 43, 1-40.

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

Dunn, L.M. & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test (PPVT4FA).* Pearson Education.

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

Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *Journal of Memory and Language*, 29, 336-360.

Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2*, 103-127.

Geraci, C., Gozzi, M., Papagno, C & Cecchetto, C. (2008). How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition*, 106, 780-804.

Haug, T. & Hintermair, M. (2003). Surveying the Need for Sign Language Tests for Hearing-impaired Children. *Das Zeichen, 64*, 220-229.

Haug, T. & Mann, W. (2008). Developing tests for sign language assessment – a review of common problems and other related issues. *Journal of Deaf Studies and Deaf Education*, 13(1), 138-147.

Herman, R. (1998a). The Need for an Assessment of Deaf Children's Signing Skills. *Deafness & Education*, 22(4), 3-8.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing British Sign Language Development: Receptive Skills Test* Forest Bookshop. Gloucestershire, UK.

Herman R, Grove N, Holmes S, Morgan G, Sutherland H & Woll B (2004) *Assessing BSL Development: Production Test (Narrative Skills)*. London: City University.

Humphries, T. & Allen, B.A. (2008). Reorganizing teacher preparation in Deaf Education, *Sign Language Studies*, 8(2), 160-180.

Jamieson, J. R. (2003). Formal and informal approaches to the language assessment of Deaf children, In B. Schick, M. Marschark, and P.E. Spencer (Eds.), *Advances in the sign language development of Deaf children* (pp. 275-288), Oxford, New York: Oxford University Press.

Kuntze, M. (1998). Literacy and Deaf children: The language question. *Topics in Language Disorders*, 18(4), 1-15.

Kyle, F., Campbell, R., Mohammed, T., & MacSweeney, M. (2007). *Developing a test of Child Speechreading*. Paper presented at the BPS Developmental Section Annual Conference, University of Plymouth, August.

Kyle, F. & Haris, M. (2006). Concurrent correlates and predictors of reading and spelling achievement in Deaf and hearing school children. *Journal of Deaf Education and Deaf Studies*, 11(3), 273-288.

Lewis, R., Vasisth, S. & van Dyke, J, (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 447-454.

Mann, W. (2008). *Facing the challenges of appropriately assessing language in Deaf children*. VDM, Germany.

Mann, W., Marshall, C. R., Mason, K., & Morgan, G. (forthcoming). The acquisition of sign language: The interplay between phonology and phonetics. *Language and Learning Development*.

Mann, W. & Prinz, P. (2006). An investigation of the need for sign language assessment In Deaf Education. *American Annals of the Deaf*, 151(3), 356-370.

Marschark, M., Lang, H.G., & Albertini, J.A. (2002). *Educating Deaf students: from research to practice*. New York, Oxford: Oxford University Press.

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

Mason, K & Rowley, K. (personal communication, December 10, 2008)

Meier, R. P. (2005). The forms of early signs: Explaining signing children's articulatory development. In B. Schick, M. Marschark, & P. Spencer (Eds.), *Advances in the sign-language development of Deaf and hard-of-hearing children* (pp. 202–230). Oxford: Oxford University Press.

Meier, R. P., Mauk, C., Cheek, A. & Moreland, C. (2008). The form of children's early signs: Iconic or motoric determinants? *Language Learning and Development*, 4, 63-98.

Meier, R. P. & Newport, E.L. (1990). Out of the hands of babes: On a possible sign advantage in language acquisition. *Language*, 66(1), 1-23.

Moores, D.F. (2001). *Educating the Deaf: Psychology, principles, and practices*. Boston: Houghton-Mifflin.

Morgan, G. (2006). Children are just lingual: The development of phonology in British Sign Language (BSL), *Lingua*, 116(10), 1507-1523.

Morgan, G., Barrett-Jones, S., & Stoneham, H. (2007). The first signs of language: phonological development in British Sign Language. *Applied Psycholinguistics*, 28, 3-22.

Prezbindowski, A.K., & Lederberg, A.R. (2003). Vocabulary of Deaf and hard of hearing children: from infancy through the preschool years. *Journal of Deaf Studies and Deaf Education*, 7, 330-345.

Schembri, A., Wigglesworth, G., Johnston, T., Leigh, G., Adam, R. & Baker, R. (2002). Issues in development of the Test Battery of Australian Sign Language Morphology and Syntax. *Journal of Deaf Studies and Deaf Education* 7(1), 18-40.

Schick, B. & Hoffmeister, R. (forthcoming). Assessing ASL skills in children: Formal Tests for Elementary School-aged Children.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*. 45A, 21-50.

Singleton, J.L. & Supalla, S.J. (2003). Assessing children's proficiency in natural signed languages. In B. Schick, M. Marschark, and P.E. Spencer (Eds.), *Advances in the sign language development of Deaf children* (pp. 289-302), Oxford, NY: Oxford University Press.

Stokoe, W.C. (1960). Sign language structure: an outline of the visual communication system of the American Deaf. *Studies in Linguistics: Occasional Paper 8*. University of Buffalo.

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

Swanwick, R. & Gregory, S. (2007). *Sign bilingual education: policy and practice*. Coleford: Douglas McLean.

Toulmin, S.E. (2003). *The use of argument* (Updated ed.). Cambridge, England: Cambridge University Press.

Table 1: Levels of complexity for the Nonsense Sign Repetition Task

		Handshape	
		simple	complex
Movement	single movement	Level 0 (10 items): Path movement *or* hand-internal movement	Level 1b (10 items): Path movement *or* hand-internal movement
	movement cluster	Level 1a (10 items): Path movement + hand-internal movement	Level 2 (10 items): Path movement + hand-internal movement

Figure 1: Structure of Assessment Use Argument (AUA)

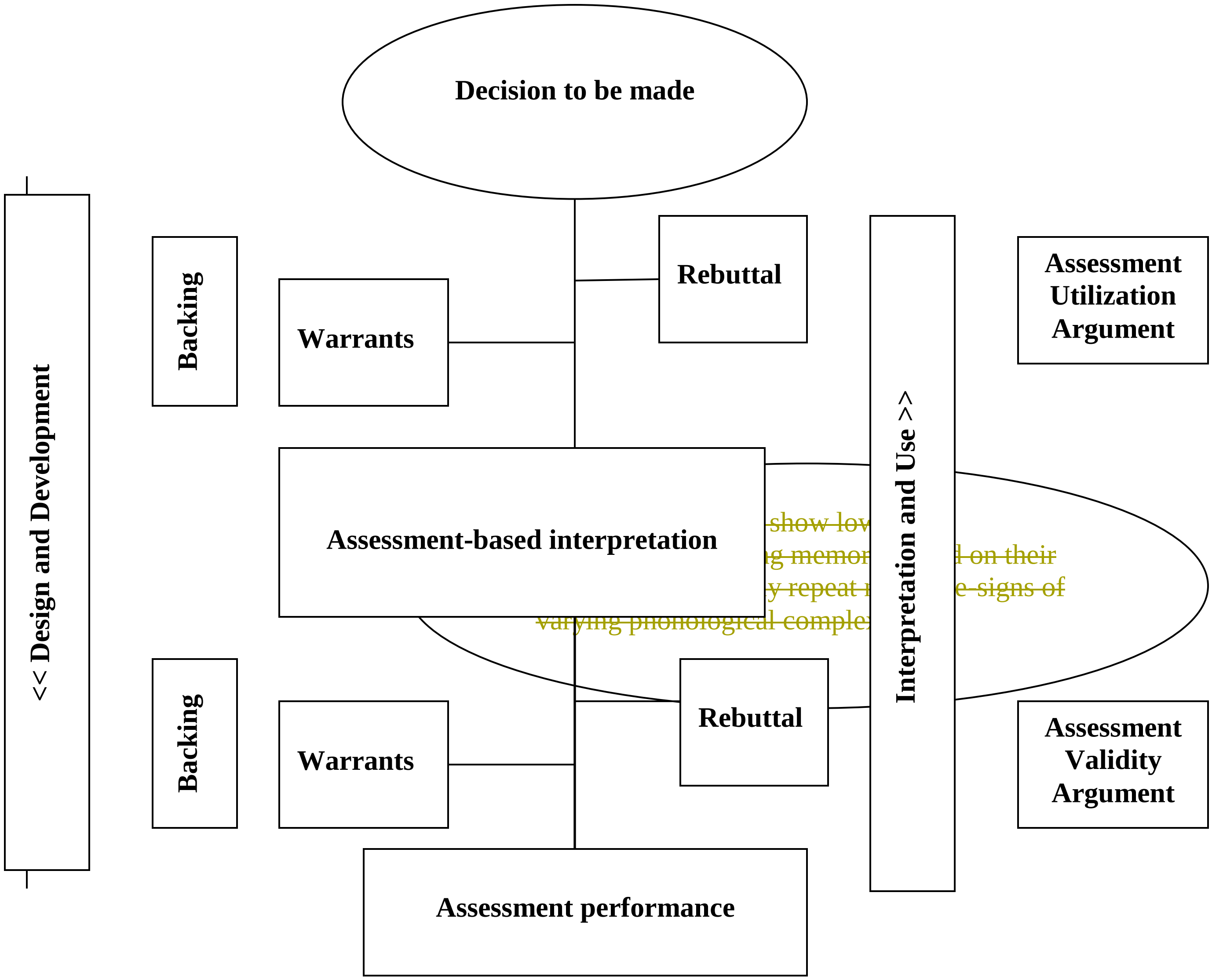

Fig. 2: Construct for measuring BSL phonological skills based on the ability to repeat nonsense signs

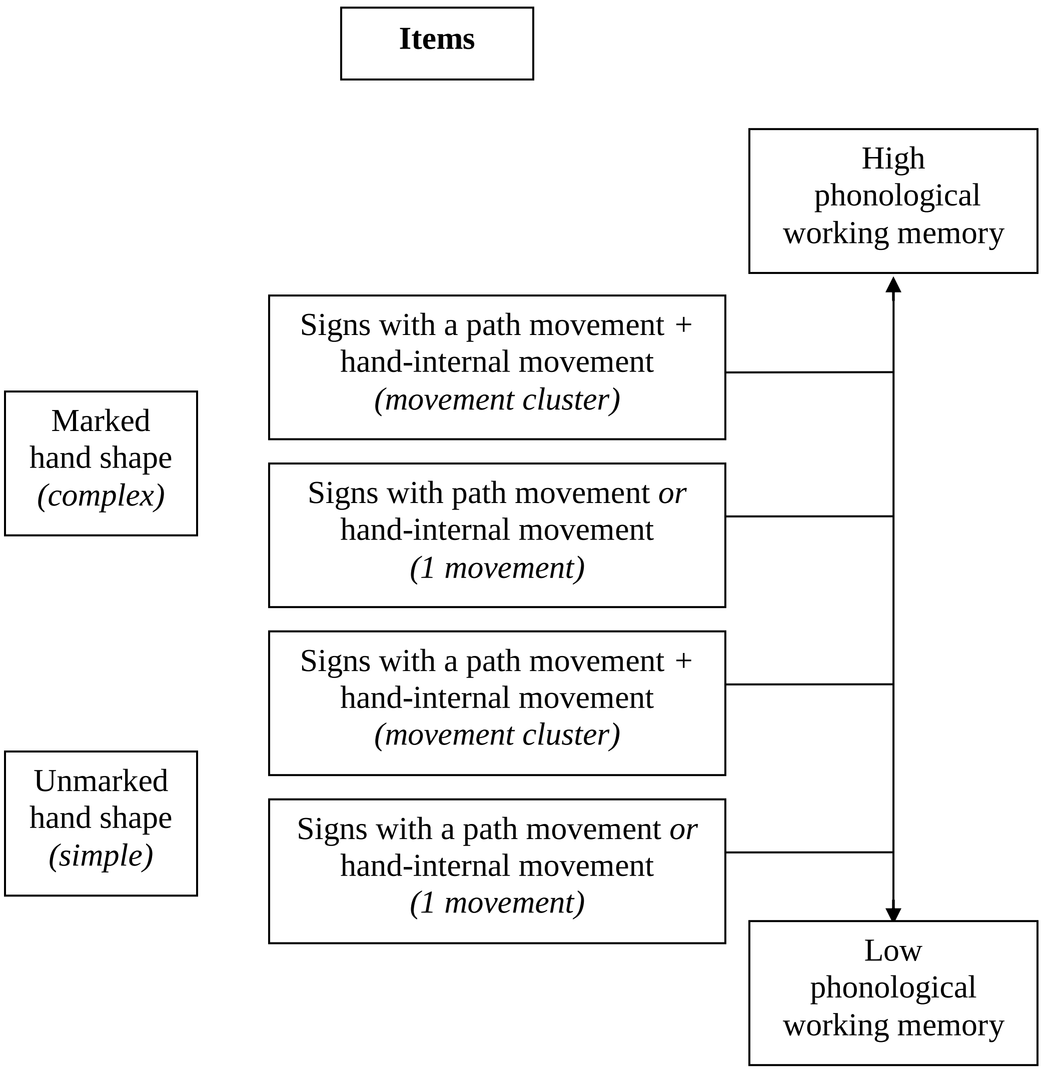

Figure 3: Structure of Assessment Use Argument (AUA) for sign language