Abstract

This study investigates children's vocabulary knowledge in Finnish Sign Language (FinSL), specifically their understanding of different form-meaning mappings by using a multi-layered assessment format originally developed for British Sign Language (BSL). The Web-based BSL-VT by Author (XXXX) was adapted for FinSL following the steps outlined by Author (XXXX) and piloted with a small group of deaf and hearing native signers (N=24). Findings showed a hierarchy of difficulty between the tasks, which is concordant with results reported previously for BSL and ASL. Additionally, the reported psychometric properties of the FinSL-VT strengthen previous claims made for BSL and ASL that the underlying construct is appropriate for use with signed languages. Results also add new insights into the adaptation process of tests from one signed language to another and show this process to be a reliable and valid way to develop assessment tools in lesser researched signed languages such as FinSL.

Vocabulary knowledge is one of the key aspects and a much-studied area within the field of children's spoken language acquisition and development - especially during their early years from 12 months to 36 months (Fenson et al. 2007; Hoff, 2003; Rowe & Leech, 2019). In comparison, research on vocabulary knowledge in school-age children is much more limited (Miralpeix, 2019). For children acquiring a signed language, there is hardly any research available on the development of vocabulary knowledge during either early or school years. Thus, this study aims to narrow the gap in our knowledge by investigating vocabulary knowledge in a population which has been rarely studied, that is Finnish signers between the ages of 4 to 15 years.

In order to produce language, children need vocabulary. This knowledge contains information on form and meaning as well as on the grammatical and usage properties of a lexical item (see the review from Yanagisawa & Webb, 2019). In that sense, vocabulary knowledge includes both *quantitative* and *qualitative* aspects. The quantitative aspect of vocabulary knowledge refers to the number of different lexical items in the vocabulary of the child (vocabulary breadth). The qualitative aspect of the vocabulary knowledge on the other hand refers to knowledge related to meaning, strength of different semantic connections, grammatical and usage properties of different lexical items (Laufer, 2013; Lin & Morrison, 2010; Qian & Lin, 2019; Schmitt et al. 2015). This knowledge, also referred to as vocabulary depth, refers to what the child knows about the different lexical items, how the different lexical items are semantically linked with each other and how the child is able to use the lexical items in different situations and interaction contexts. Vocabulary knowledge is regarded as the foundation of all language use and learning and also as an important predictor for later language and literacy skills both in spoken and signed language (Laufer, 2013; Lin & Morrison, 2010; Qian & Lin, 2019; Schmitt et al. 2015). Consequently, assessing vocabulary knowledge is instrumental for describing and evaluating children's language skills and monitoring their developmental trajectories. However, contrarily to spoken languages there is a considerable shortage of assessment tools designed to specifically evaluate vocabulary knowledge in signed

languages and little is known about how children acquire quantitative and qualitative aspects of vocabulary knowledge. The assessment of vocabulary knowledge can provide essential information on children's language skills and development which is particularly useful for parents, practitioners and clinicians working with the child.

Most of our existing knowledge about vocabulary development of children acquiring signed language has been informed by studies focusing on American Sign Language (ASL) or British Sign Language (BSL). In this study we investigated vocabulary knowledge in two lesser researched populations of signers: deaf and hearing children acquiring Finnish Sign Language (FinSL) from birth. Due to the lack of any existing tests for FinSL, including vocabulary assessments, we adapted one of the few tests specifically developed to measure vocabulary knowledge in a signed language, the British Sign Language Vocabulary Test (BSL-VT, Author, XXXX) for FinSL, following the steps outlined by Enns and Herman (2011) and Author (XXXX). We begin by providing some background on vocabulary knowledge and signed vocabulary acquisition, describing similarities and differences compared to spoken language vocabulary within the context of vocabulary assessment. We then outline the process of adapting the BSL-VT for FinSL and present findings from a small pilot with hearing and deaf Finnish signers between the ages of 4 to 15 years. In our discussion of the results we draw a comparison to previous studies that have investigated children's vocabulary knowledge in two other signed languages, ASL and BSL. In doing so, this paper makes an important contribution to the field by narrowing the gaps in our knowledge about lesser studied signed languages in general and, more specifically, regarding the development and assessment of signed vocabulary development among children.

Aspects of vocabulary knowledge

From their first vocabulary spurt around 19 months children's vocabulary development accelerates and keeps growing (Fenson et al., 2007; Nation & Waring, 1997). During this process children acquire both the quantitative and qualitative aspects of their vocabulary knowledge. As the number of different lexical items increases in the child's vocabulary the child also starts forming semantic connections, grammatical and usage properties between the different lexical items. By acquiring new lexical items and building more complex and organized semantic networks to structure their growing vocabulary, children create an initial mapping between phonological form and meaning of new lexical items acquired (Clark, 2009). Thus, to maintain the accelerated vocabulary growth, children organize newly acquired items (spoken or signed) in a structured manner in their mental lexicon. This includes knowledge of an item's form and meaning as well as its grammatical functions and usage properties (e.g., collocations). The phonological form and meaning connections are established through links, or mappings, between lexical items, resulting in multiple networks. In these networks the phonological form of a lexical item gives the child access to its meaning and, at the same time, meaning provides access to the phonological form. For example, when acquiring the word *apple* a child might for example initially draw on information related to shape, e.g., a small, round object that can be eaten, to aid the meaning of this particular phonological form a-p-p-l-e [æp(ə)l] and later start adding different semantic and categorical links to other new or existing words in the lexicon. These links may be of taxonomic (e.g., *banana, orange*), semantic (e.g., *fruit, food*) or thematic (e.g., *juice, red, skin*) nature. Laufer et al. (2004) regard this form-meaning mapping network as the *strength of vocabulary knowledge*.

As the child's vocabulary grows more and more links are generated, initially through mappings between the form and meaning of lexical items, which then form larger semantic networks between the lexical items acquired (Clark, 2009). Consequentially, capturing the complex nature of vocabulary knowledge requires attention to what and how many different lexical items the child knows, what the child knows about the semantic connections of different lexical items and

how appropriately and effectively the child uses the lexical items in different contexts. Yanagisawa & Webb (2019) describe vocabulary as a multifaceted phenomenon that contains different dimensions of knowledge rather than being one unitary ability. The different dimensions of vocabulary knowledge represent an essential part of children's linguistic abilities and literacy skills (Laufer, 2013; Lin & Morrison, 2010; Qian & Lin, 2019; Schmitt et al., 2015). Children with broad and deep vocabulary knowledge have been found to be stronger readers, have better reading comprehension, be able to learn new words more quickly and use words more flexibly across a variety of contexts compared with their age-peers with lower vocabulary knowledge (Dickinson & Porche, 2011; Proctor et al., 2012; Munson et al., 2005). As a result, vocabulary knowledge serves as a highly important stepping stone for further language and literacy skills. For this reason, assessing the different dimensions of vocabulary knowledge is important to gain better understanding of child's language abilities.

Traditionally, most of the available test instruments assess vocabulary knowledge in spoken language, e.g., Peabody Picture Vocabulary Test, (PPVT, Dunn & Dunn, 1997); Expressive One Word Picture Vocabulary Task (EOWPVT, Brownell, 2000); Receptive One Word Picture Vocabulary Task (ROWPVT); British Picture Vocabulary Scale (BPVS, Dunn et al., 1997). These instruments generally focus on measuring the quantitative aspect of vocabulary knowledge, i.e., size, rather than qualitative aspects of vocabulary knowledge e.g., the *strength of the vocabulary knowledge*. Additionally, existing vocabulary tests tend to concentrate on one mapping between the form and meaning, for instance, by asking the test taker to match a picture response (meaning) to the phonological form of a target word or sign (form) (e.g., the *Peabody Picture Vocabulary Test*, *PPVT*, Dunn & Dunn, 1997) or to produce a phonological form to match the meaning represented by a picture prompt (meaning -> form mapping) (e.g., the *Expressive One Word Picture Vocabulary Test*, *EOWPVT*, Brownell, 2000). While these approaches might be considered as providing a more representative picture of the learners' vocabulary breadth compared to tests that

focus on vocabulary depth, using fewer items, they have been questioned by some researchers as insufficient (Hadley & Dickinson, 2020). Although there is a growing number of voices calling for test instruments that assess both quantitative and qualitative aspects of vocabulary knowledge (Yanagisawa & Webb, 2019) such instruments are still rare, specifically for use with children acquiring a signed language.

Vocabulary development in signed language

Compared to vocabulary acquisition and the assessment of vocabulary in spoken languages, much less is known about these processes in signed languages. The study of signed vocabulary development is interesting for a number of reasons including the sociolinguistic complexities and heterogeneity of the signing population. There are four groups of signing children: 1) deaf children who have deaf parents, 2) hearing children of deaf parents (also referred as KODA, Kids of Deaf Adults), 3) deaf children who have hearing parents, and 4) hearing siblings of deaf children. It is well-documented in the literature that those deaf children growing up in deaf families acquire signed language within a native language learning and reach the same developmental milestones at a similar pace to hearing children (Anderson & Reilly, 2002; Lillo-Martin, 1994). The same applies to hearing children with deaf parents who are exposed to both signed and spoken language. However, the difference is that these children often grow up bilingually (access to signed and spoken language from birth) and can be regarded as heritage signers (Chen Pichler et al. 2018). Thus, recent research that focused on the bilingual context in which these children learn has found unique effects on the developmental patterns of each of their languages as a result of their bimodal status (see the review in Chen Pichler et al. 2018; Reynolds, 2018). In contrast, the majority (i.e.,

90-95%, Mitchell & Karchmer, 2004) of deaf children and their hearing siblings with hearing parents grow up in widely differing language backgrounds. In case of deaf children, this can have an effect on their age and level of signed language exposure (Mayberry & Eichen, 1991; Mayberry, 1993, 1994) as well as on both the quantity (Marschark, 1997) and quality of their signed language input (Singleton & Newport, 2004). Consequently, children from this group tend to acquire vocabulary at a slower pace and have an overall smaller signed lexicon size (Anderson, 2006; Lederberg & Spencer, 2009). One reason for this may be children's more limited incidental exposure to language in the form of signed interactions between other family members (Lu et al., 2016). Contrarily to hearing families with a shared language, hearing parents who decide to learn a signed language to communicate with their deaf child tend to use it exclusively when addressing their child rather than for communicating with hearing family members (Marschark, 1997).

In addition to the sociolinguistic circumstances which affect signed vocabulary development, there are other ways in which signed vocabulary development is different from spoken language. Some of these include the lack of a written form for signed languages, the distinction between a native and a non-native lexicon (fingerspelt representations of spoken words) or the more profound effect of iconicity (i.e., resemblance between form of the sign and its referent) on items in the signed lexicon, e.g., DRINK, SLEEP. However, despite these differences, vocabulary acquisition in signed languages appears to be very similar to that of spoken languages. For instance, Anderson and Reilly (2002), who adapted the MacArthur-Bates Communicative Developmental Inventory (CDI) into signed language, reported that deaf children's acquisition of American Sign Language (ASL) vocabulary within particular semantic domains, such as question words, emotion signs, and cognitive verbs, was comparable to that found in hearing peers.

This study focuses on deaf children with deaf parents and hearing children with deaf parents in an effort to validate the newly adapted FinSL-VT and also to establish a trajectory of vocabulary development in native users of FinSL.

Assessing vocabulary knowledge in signed language

Perhaps unsurprisingly the number of available tests for signed languages is limited when compared to spoken language. This number is even smaller for tests that were specifically developed to measure signers' vocabulary knowledge. One of the few available instruments is the MacArthur Communicative Development Inventory (MacArthur CDI) (Fenson et al. 2007), a parental report form that measures early vocabulary development in children between eight to thirty-six months. The MacArthur CDI has been adapted for ASL (Anderson & Reilly, 2002) and BSL (Woolfe et al., 2010). Prior to the BSL-VT the only vocabulary test available for older children was the PERLESKO (Prüfverfahren zur Erfassung Lexikalisch-Semantischer Kompetenz) by Bizer and Karl (2002a), which assesses receptive vocabulary in three language modalities: German Sign Language (DGS), spoken German and Written German in children aged seven and thirteen years (Bizer & Karl, 2002b). There are different reasons for the reported shortage of available test instruments for signed languages: one is the early stage of signed language research, which has mostly been focusing on American Sign Language (ASL) and some European Signed Languages, including British Sign Language and German Sign Language. Other reasons include the small size and make-up of the signing population which is known to be very heterogeneous with regard to the degree and type of hearing loss, age of diagnosis, type of amplification, and language exposure at home and in school, and many other variables (Allen, 2014). It is worthwhile mentioning that all of the existing signed language vocabulary test measure vocabulary breadth with the primary aim to assess the number of signs known by a test taker rather than the level, or depth of that knowledge.

Assessing vocabulary knowledge is a quite complex task/process and largely depends on how one defines what 'knowing a word' means (for a conceptual review of measuring vocabulary knowledge in spoken language, see Hadley & Dickinson, 2020). As mentioned before, vocabulary knowledge does not rely on one type of information (qualitative, quantitative) only but

rather a combination of these aspects. Children's level (or depth) of vocabulary knowledge largely depends on the number of encounters with a particular word as well as the context in which these multiple encounters took place. Both the quality as well as quantity of such encounters may vary across children, depending on a number of variables, e.g., SES (Hoff, 2003). Similarly to the way in which children gradually build their vocabulary knowledge through multiple encounters, they draw on different types of information to acquire the meaning of words (or signs). This may include perceptual information (e.g., seeing a tree while hearing the word 'tree'), linguistic information (e.g., reading about the word 'galaxy' in a dictionary) or a combination of both. Wauters and colleagues (2003, 2008) refer to this as *Mode of Acquisition* (MOA, Wauters et al., 2008, p.175). MoA draws on the idea that knowledge of a word's meaning can be understood as the interrelated array of the many different associations one has with that word, both linguistic associations but also expressions and perceptual associations. Wauters and colleagues found that MoA can be reliably measured by asking adults or children to judge the MoA of words (Wauters et al., 2003, Wauters et al., 2008) on a five-point scale. We adapted the MoA scale for use with signing children to validate the target items for the FinSL-VT.

Notably, signed languages have always had a minority status compared to spoken languages, which traditionally contributes to the multilingualism of signed language users. To successfully fit in the education system, working life and the wider linguistic society requires the signer to have knowledge of the spoken/written language of the linguistic majority. For this reason, elements of multilingualism and multiculturalism can be observed in signing adults and children, who use different modalities (such as signing, writing and speaking) as part of their daily communication. Consequentially, the initial setting and context for language acquisition of deaf and hearing children acquiring sign language(s) are at least bilingual but often also multilingual. This makes it highly challenging to develop and standardize language assessment tools for the small and heterogenous population that is children acquiring a signed language. At the same time, the

considerable variation with regard to early language experience of signing bilinguals makes it particularly important to assess both quantitative and qualitative aspects of children's vocabulary knowledge. One of the few instruments developed for signed language which allows this dual focus is the web-based British Sign Language Vocabulary Test (BSL-VT, Author, XXXX).

The web-based BSL Vocabulary Test (BSL-VT) is based on a theoretical model from second language acquisition which proposes that the relationship between word form and word meaning can be expressed in four degrees of strength (Laufer et al., 2004) namely active recognition *(choose the target word from four options)*, passive recognition *(choose the meaning of the target word from four options)*, active recall (supply the L2 word) and passive recall. This terminology and the format of the tasks were adapted for use with signed languages, resulting in the following changes: *meaning recognition (matching a sign to four pictures)*, *form recognition (matching a picture to signs)*, *form recall (picture naming)*, and *meaning recall (sign association)* (see Figure 1). The web-based BSL-VT was developed for children aged 4-15 years to examine variation in deaf children's understanding of these four different mappings between phonological form and meaning in single BSL signs (Author, XXXX). It consists of 120 vocabulary items (12 sets of 10 items developed for ages 4 to 15 years). The item selection process was informed by various sources, including a BSL norming study (Vinson et al., 2008), a receptive vocabulary test for DGS (PERLESKO; Bizer & Karl, 2002b), a number of standardized English vocabulary tests, as well as feedback provided by a group of deaf and hearing researchers and teachers of the deaf, all of whom were fluent signers. Items are arranged in their order of difficulty, based on their position in any of the other vocabulary assessments (where available) and on judgment by the deaf and hearing teachers (Author, XXXX).

Two of the four BSL-VT tasks (see Figure 1), *meaning recognition* and *form recognition*, assess vocabulary comprehension: In the *meaning recognition* task, children see a sign and need to select the image that best illustrates the meaning of the sign. This format is the same as

in some standardised tests for spoken language, e.g., PPVT, BPVS. In the *form recognition* task, children are presented with a stimulus image and four signs and have to select the sign that best matches the image. The remaining two tasks, *form recall* and *meaning recall*, assess vocabulary production. In the *form recall* task, children see an image on the computer screen and are asked to produce the corresponding sign. Finally, for the *meaning recall* task, participants see the target sign presented on their screen and have to supply three different signs with an associated meaning. Figure 1 illustrates the construct of strength of form-meaning mapping, ranging from weak to strong understanding of form-meaning mapping.

>>> Insert Figure 1 here <<<

Figure 1. FinSL-VT tasks (2019).

The same 120 items are used across all four tasks of the BSL-VT. This multi-layered approach enables a more in-depth evaluation of vocabulary knowledge as the test evaluates both quantitative and qualitative aspects of vocabulary knowledge, that is, vocabulary size and strength. An example illustrates this (see Figure 1): a child, who is not able to produce the sign for BIRD when prompted with a picture showing a bird (form recall task) might be able to produce another, meaning-related sign, e.g., EAGLE when presented with the target sign (meaning recall task), or recognize its meaning when presented with the sign for BIRD and a set of picture responses (meaning recognition task) (Author, XXXX, pp. 95). As a result, the BSL-VT recognises that the child holds some knowledge of the target word/sign whereas - on any of the abovementioned standardized vocabulary tests - the same child's response would have been scored as 'fail'. This makes it a particularly valuable tool to assess children from linguistic minority backgrounds, including deaf children, who tend to perform low on standardised language assessments (Hadley & Dickinson, 2020).

The BSL-VT has already been adapted for another signed language, that is American Sign Language (ASL) (Author, XXXX) although for a smaller age range (6–10 years). As part of this process two deaf native signers with teaching backgrounds reviewed the list of 80 items from the BSL vocabulary test and identified those that were appropriate for use in ASL. As a result of these discussions, 66 of the 80 items were accepted for adaptation without further changes whereas 11 items required a change to the target item (and development of new items) and three items required a change to the label, due to differences between British English and American English (Author, XXXX).

Research that compared children's understanding of different form-meaning mappings in ASL and in BSL found statistically significant differences in children's performances across the four tasks (Author, XXXX; Author, XXXX) for both signed languages. According to these findings signers who successfully recalled the meaning or form of a sign tended to also recognized that same sign's form or meaning. The results from research on the BSL-VT and the ASL-VT both support the claim by Author (XXXX) that there is a hierarchy of strength of signed vocabulary knowledge which suggests that knowing a sign is a gradual (rather than an instantaneous) process which involves different levels of knowledge.

Assessing vocabulary knowledge in Finnish Sign Language

This study adds to previous research that has investigated vocabulary knowledge of signing deaf children by extending the focus to a lesser researched signed language (FinSL), and population (Finnish signers). Only a few studies have focused on FinSL acquisition and development among signing children (Takkinen, 2005) and only one study investigated early development of vocabulary with a focus on young KODAs between the ages 12 and 30 months (Author, XXXX). One of the reasons for this limited research is the lack of assessment tools for FinSL, specifically for use with children. Although signed language research in general is still a

young field that only started in the late 1950s in the U.S. (Stokoe, 1960), FinSL has been studied considerably less in comparison to other signed languages, in particular ASL or BSL. Over the course of the last years there has been growing research interest in FinSL structure (see e.g., Jantunen, 2017a; 2017b; Puupponen, 2019) coupled with the release of the Corpus of Finland's Sign Languages in 2019. The current lack of developmental patterns in young Finnish signers' language acquisition has made it difficult to provide the linguistic foundation necessary for developing language assessment tools. These limitations and the need for a vocabulary test triggered the decision to adapt an existing test from one of the more researched signed languages, namely BSL. In the following section we provide a detailed description of the steps that were involved in adapting the web-based BSL-VT for FinSL. The following research questions were specified:

1) What linguistic and cultural differences exist between BSL and FinSL and how can these differences be addressed when adapting the BSL-VT for FinSL?

2) How are the psychometric properties, including reliability and validity, of the FinSL-VT assessed?

3) Does Finnish children's performance on the FinSL-VT indicate a similar hierarchy of strength of vocabulary knowledge as reported for BSL and ASL?

Methods

Adaptation process

In order to assess signing children's vocabulary knowledge, specifically their understanding of different form-meaning mappings in FinSL, the Web-based BSL-VT (Author,

XXXX) was adapted for FinSL. The adaptation process was conducted by following the stages outlined by Enns and Herman (2011) and also by Author (XXXX). These stages are presented in Table 1. Each stage will be described in more detail in the following sections.

Table 1. Stages for adaptation process from BSL-VT to FinSL-VT based on Enns and Herman (2011) and Author (XXXX)

>>> Insert Table 1 here <<<

Stage 1. The first step of the adaptation process was setting up a panel of FinSL experts. This panel consisted of six members, four of whom were deaf and two hearing. Five panel members were native signers and the sixth was a hearing, fluent L2 signer, who had learnt FinSL during adulthood. All members had previous experience working with deaf and hearing children acquiring signed language e.g., at schools. The panel worked closely with the first author during stages 2 and 3. Additionally, all panel members participated in the data collection during stages 7 and 8 and provided feedback on drafted test materials during stage 9.

Stages 2 and 3. During these stages two of the deaf native panel members reviewed the list with all 120 BSL-VT items and discussed whether they were appropriate for use in the FinSL version. Where needed, the panel suggested alternative items. The new item list was then discussed with all panel members. From the original 120 target items of the BSL-VT the panel decided to change a total of 41 items (34%). In 36 cases this decision was made due to the lack of a compatible sign in FinSL (e.g., the FinSL equivalent was a compound sign). For example, in FinSL the sign TEACHER is produced as TEACH PERSON; similarly, the sign UNIVERSITY is produced as COLLEGE INSTITUTION. For other target items it was challenging to find suitable

distractor items. Finally, all 120 vocabulary items in FinSL version was decided and the same 120 items were used across all four tasks.

The meaning recognition and form recognition tasks use a multiple-choice format with four responses, including the target and three distractors. These distractors were phonologically, semantically, or visually related to the target item. For the phonological distractor a sign was chosen that shared one or more similar phonological parameters (e.g., place, orientation, location, movement) with the target (e.g., in Figure 1 SPIDER). For the semantic distractor a sign was chosen that had a semantic link to the target item (e.g., in Figure 1 BUTTERFLY). The visual distractor was either an unrelated sign or a gesture that resembled the shape of the sign prompt or had a similar action presented in the target picture (e.g., in Figure 1 AIRPLANE). As the distractors needed to be related to target items phonologically and visually and there were only 17 phonologically identical signs in both BSL and FinSL, new phonological and visual distractors had to be developed for the target items. For instance, the target sign DOG in BSL (signed with two n-shaped hands repeatedly pointing downwards) is signed in FinSL by tapping the side of the thigh with the 5-shaped hand. However, this location is used rather infrequently for signs. Five items were replaced due to cultural differences (e.g., the sign RUGBY used in the BSL-VT was changed to FOOTBALL in the FinSL version) as was one item (DVD) because the sign is fingerspelled in FinSL.

Stage 4. After the 120 target items had been agreed on by the panel, the same two signers that reviewed the target and distractor signs during Stage 2 video-recorded all final signs for the FinSL-VT in high definition format. One of them also acted as sign model for the general test introduction and the instructions for each task. All picture prompts and distractors were created by a deaf illustrator. This person, a native signer himself, met with the first author to discuss the multi-task format of the test and the purpose of each picture.

Stage 5. The validation of target items was done in two stages: one focusing on target signs and one with focus on target pictures. Target signs were validated based on teacher ratings of the type of knowledge necessary for acquiring each target sign ('Mode of Acquisition'), an approach that was also used by Author (XXXX), based on the aforementioned work done by Wauters and colleagues (2007). The Likert-type scales that teachers used ranged from '1' (perceptual associations/learning the meaning through experiences) to '5' (linguistic associations/learning the meaning through language). Similar to the study by Author (XXXX), four Finnish teachers of the deaf (two deaf and two hearing) made use of the full range of ratings suggesting that the item pool for the FinSL-VT was appropriate. MoA ratings showed an average score of 3.3 (1=12%, 2= 8%, 3=23%, 4=25%, 5=23%). Overall the Finnish teachers gave slightly higher ratings compared with the results of Author (XXXX), which could be due to the considerable number of signs that had been replaced as part of the adaptation. In order to validate the target pictures for the FinSL-VT, deaf (N=4) and hearing (N=3) students from the Master Program of Sign Language at XXXX were presented with pictures of all target items and asked to write down for each their three best guesses what it meant. In those cases where 50% or more of the students did not guess the correct answer the picture (not the item) was replaced. This happened in ten cases (out of 120).

Stage 6. The same web-based format was used as for the BSL (Author, XXXX) and ASL (Author, XXXX) versions. The meaning recall task was set up as a repeated meaning association task, in accordance with a change that was introduced for the ASL-VT in an effort to gain more detailed information about the way children organise their mental lexicons (Sheng et al., 2012; Author, XXXX). Consequently, the task instructions were modified and children encouraged to think about three signs (rather than one) with related meaning to the signed prompt. At the end of this stage, the prototype of the FinSL-VT was completed with four different tasks (meaning recognition, form recognition, form recall and meaning recall) that contained 120 test items each.

Stage 7. A pre-pilot phase was carried out during which two deaf adults completed all four tasks and provided feedback related to items and also regarding the web-based format. Both participants scored close to 100% on three out of four tasks (Meaning Recognition 98%, 89%; Form Recognition 95%, 92%, Form Recall 91%, 86%). As expected due to the nature of this task, scores were notably lower for the Meaning recall task (63%, 62%). This was the same for the ASL version (Author, XXXX). As an additional validation mechanism, the two comprehension tasks (Form Recognition, Meaning Recognition) were presented to a control group of 24 age-matched hearing children (M: 10;0 years SD: 3.3 years, range 4;10–15;11 years; 18 females) with no previous knowledge of FinSL. The percentage of correct responses on the Meaning Recognition task was 37% (SD: 9.1, Range: 30–61) and on Form Recognition 36% (SD: 9.0, Range: 27–63). The results for both tasks are above chance level (25%). One possible reason for this is that the signer mouthed the Finnish word for some signs and the hearing children were able to read their lips. Another reason may be the iconicity (resemblance to action, movements, location and shapes of object and/or person) of some signs which may have aided non-signers in selecting the correct response. This has been discussed in previous studies e.g., by Hermans et al. (2010).

Stage 8. Pilot testing was carried out with a total of 24 children including nine deaf and fifteen hearing signers (18 females) between the ages 4;1 to 15;7 years (M: 9;11 SD: 3.6) (two children in each age-group) acquiring FinSL from their deaf parent(s) (see Table 2 for demographic information on the participants). All participants had at least one deaf parent. The children were recruited through the author's existing links to the Finnish signing community. All children were native users of FinSL, which in the context of this study means that they had exposure from birth. As mentioned before, many signing children grow up in a bilingual or even multilingual language environment. In Finland, in addition to two national languages (Finnish and Swedish), Saami, Romani, Finnish Sign Language (FinSL) and Finland-Swedish Sign Language (FinSSL) are recognized as minority languages in Finnish legislation. In case of this study, all participating

children were acquiring at least two languages (see Table 2). All parents described their children as typically developing with good or excellent skills in FinSL in comparison to same aged peers. This information was provided in form of a questionnaire on children's linguistic environment and demographic background. Equally, no concerns were expressed by parents regarding children's cognitive or linguistic performance. All parents had completed at least secondary level of education (either upper secondary school or vocational school) and 12 out of 48 parents had a university degree. Ethical approval for the study was obtained from the Ethical Committee at the XXXX prior to collecting any data.

Table 2. Demographic information of participants*.

>>> Insert Table 2 here <<<

Additional data collected in the pilot stage. Participants between the ages of 4 and 11 years (N=15) also completed the web-based FinSL Receptive Skills Test (FinSL-RST, Author, XXXX) that assesses children's understanding of morphological and syntactical structures in FinSL. It has been developed for signing children between the ages of 3 and 11 years. The FinSL-RST is an adaptation of the BSL Receptive Skills Test (Herman et al., 1998). Like the BSL-RST the FinSL-RST uses a multiple-choice format where children see a signed prompt on their computer screen next to four images and select the one that matches the meaning. The test consists of 47 items and takes about 15 minutes to complete. This test was included to provide an additional measure of children's signed language proficiency.

As children acquiring signed language represent a highly heterogeneous group with particular regard to their hearing status, degree of hearing loss and their access to language in general, demographic information provides necessary detail on the context in which language is

acquired. For this reason, background information was collected from all participants by means of a parental questionnaire with a focus on children's language context at home and in school. The questionnaire consisted of items adapted from three existing instruments, namely the PaBiQ (Questionnaire of Parents of Bilingual Children in Tuller, 2015), the BiLEC (Bilingual Language Experience Calculator in Unsworth 2013) and the MAIN (Multilingual Assessment Instrument of Narrative in Gagarina et al., 2012), all of which have been used in previous research with bilingual and multilingual children (Rodina, 2017; Tuller, et al. 2018). Items included questions on children's age, hearing status of the child and parents, language usage practices at home and outside the home context, and the amount of language exposure from different languages. However, this information was not used as a selection criterion for the study. Thus, no child was excluded from the study based on the information provided in the questionnaire.

Procedures

All FinSL-VT tasks were presented to each child individually either in their home, the day care centre or at school by a deaf or hearing test administrator (from the panel of experts), who was a fluent signer. Deaf panel members administrated the test with all the hearing participants and both hearing and deaf panel members administrated the test with the deaf participants. Because all panel members had been working actively with the test material during the test adaptation process they were highly familiar with the material as well as the administration procedures.

All four FinSL-VT tasks were administered individually following the same order (form recall -> form recognition -> meaning recall - > meaning recognition) recommended by Author (XXXX) in two to four separate sessions depending on the child's age and motivation. The reason for using this order was to minimise learning effects since all the tasks use the same 120 vocabulary items. In addition to the FinSL-VT tasks a smaller subset (N=14) of the children also completed the FinSL-Receptive Skills Test (RST). The children completed all the four tasks of

FinSL-VT and FinSL-RST in two to four sessions that each lasted 45-60 minutes. All the tasks were administered on the internet by the test administrator. The two recognition tasks used a multiple-choice format and were scored as '0' for an incorrect response or '1' for a correct response. Answers were automatically scored and saved onto a web-server at the first author's university. For the two recall tasks the children's answers were video-recorded and entered by the test administrator in a text box provided on the screen next to the prompt by using Finnish glosses. These text entries were automatically saved onto the same web-server. They were scored as either '0', '0.5', or '1'. This was done by the first author following test administration. It was a change in protocol from previous studies (e.g., Author, XXXX), which used live scoring, in order to further strengthen the reliability of scoring. Inter-rater reliability on the scoring of the two recall tasks was carried out by a deaf native signer who scored the responses of four children (20% of the sample) which were then compared to the scores prepared by the first author. Cohen's kappa was calculated to determine the level of agreement between raters' judgements. Following the guidelines provided by Landis and Koch (1977), agreement of judgments for scoring the form recall task (k = .89) and for scoring the meaning recall task (k = .81) were very good.

The form recall task response was coded as correct and scored as '1' if the child produced the expected FinSL sign to name the target item. Responses were coded as partially corrected and scored as 0.5, if the child produced a sign that was outside the immediate range of expected answers but still suggested that the child knew the meaning of the target (e.g., when the target sign was FRIEND and the child answered a MAN the answer was scored 0.5). For the meaning recall task responses were scored as '1' if the child produced a categorical response to a target item (e.g., a synonym *happy–excited,* antonym *strong–weak,* coordinate *cherry–strawberry*, subordinate *bird–swan*, superordinate *mouse–animal*, direct negation of the stimulus sign *forget–remember*). Any responses that were semantically related with the target item (e.g., *hospital–doctor, mouse–small*) received a '0.5' score. The number of expected responses was increased from one to

three. This change was done in alignment with the ASL-version of the test in order to gain a more in-depth understanding of children's semantic knowledge (Author, XXXX; Sheng et al., 2012). Such understanding is reached by measuring both storage and accessibility of different types of semantic relations (Sheng & McGregor, 2010). Accordingly, the maximum score for the meaning recall task was 360 and 120 for the other three tasks.

Stage 9. Following the collection of pilot data the expert panel reviewed the test and distractor items from all four tasks. To do so panel members used a 3-point rating scale to indicate the need for major or minor revision of an item. Major revisions included replacing target signs or pictures. Minor revisions included re-filming signs, making small changes to target pictures and/or distractors. Most of the suggested revisions were related to the pictures; only five target signs needed to be re-filmed. Based on the rating and comments of the panel, changes were made for a total of 27 items (21 minor and 6 major revisions) out of 120.

Results

Participants' performance scores on the four FinSL-VT tasks are presented in Table 3. The raw scores show a hierarchy in children's performance, with the meaning recognition task showing the highest performances and meaning recall task the lowest performance. The scores of the meaning recall tasks were considerably lower compared with other tasks. In addition, children's performance on this task varied more notably compared with other three tasks. In order to provide evidence in support of reliability and validity of the FinSL-VT the internal consistency, item analysis, inter-rater reliability, content validity, construct validity and concurrent validity were explored. These psychometric properties will be presented in the next section.

Table 3. Raw scores for the FinSL-VT tasks and FinSL-RST task

>>> Insert Table 3 here <<<

Reliability

Internal consistency was assessed using Cronbach's alpha and corrected item-total correlations. Cronbach's alpha for all tasks together was .717 which is above the .700 that is considered to be the lower limit for an adequate coefficient (Nunnally, 1978). Corrected item-total correlations ranged from .790 to .862 which is considerably above the .30 that is regarded the limit for corrected item-total correlations (Nunnally & Bernstein, 1994). The meaning recall task indicated the lowest corrected item-total correlation (.790) and a high Cronbach's alpha for any item deleted (.960) compared with other tasks. For this reason, Cronbach's alpha was measured again but this time separately for each task. The average alphas of three tasks, namely meaning recognition (.800), form recognition (.790) and form recall (.790) met the requirements of coefficient, whereas the significance for the meaning recall task was low (.500). As shown in Table 3 the children scored considerably lower on the meaning recall task compared with the other three tasks which affected the results in internal consistency. The possible reasons for these lower scores will be addressed in the discussion of this paper.

An *item analysis* was run to identify any items that had been answered correctly or incorrectly by all participants and should be removed for this reason. The analysis showed no such items. All participants correctly answered seven items on three of the four tasks (*meaning recognition*, *form recognition* and *form recall*) but none of these items were passed by all participants in the meaning recall task. Additionally, no item was failed by all participants within one task. Thus, no item needed to be removed.

Inter-rater reliability was evaluated for the two production tasks, i.e., form recall and meaning recall tasks. The first author coded and scored the responses of all the participants. Afterwards, the responses of four participants (20% of the sample) were coded by an independent

deaf research assistant that was a native user of FinSL. Cohen's kappa was calculated to determine the level of agreement between the raters' judgements. Following the guidelines provided by Landis and Koch (1977), the agreement between raters' judgements for scoring the form recall task (k = .890) and for scoring the meaning recall task (k = .810) were very good. Items from the two recognition tasks were automatically scored by the computer upon selection of the responses via mouse click.

Validity

Content validity of test materials was ensured in three different steps: 1. by working closely with the deaf/hearing expert panel during the whole adaptation process, 2. by collecting MoA ratings from teachers of deaf children (see stage 5 of the adaptation process) and 3. by involving seven students of the Master Program of Sign Language at XXXX to validate the target pictures (see stage 5 of adaptation process). All these stages were important in ensuring the content validity of the FinSL-VT even more so because of the larger number of test items and distractor items that had to be revised compared to the adaptation process between the BSL-VT and ASL-VT (see stage 2) (Author, XXXX).

Construct validity was evaluated by examining whether participants' performance on the four tasks correlated with age. Additionally, the differences between participants' performance across the four different FinSL tasks were investigated. This was done to test whether the observed hierarchy of task difficulty would meet our predictions based on the underlying model of strength of form-meaning mapping. First, we carried out bivariate correlations between each of the tasks and age, using a Pearson correlation coefficient, with the alpha level reduced to .013 to compensate for multiple (k=4) comparisons (see Table 4). Findings revealed strong, positive correlations which were statistically significant between age and meaning recognition, $R(24) = 0.798$, $p < .01$, age and form recognition, $R(24) = .774$, $p < .01$, age and form recall, $R(24) = .777$, $p < .01$ and age and

meaning recall, R(24) = .764, $p < .01$. Next, we carried out partial correlations between the different tasks, controlling for age, with the alpha level reduced to .008 for multiple (k=6) comparison. Performances on all tasks remained significantly correlated: meaning recognition and form recognition, R(21) = .860, $p < .001$, meaning recognition and form recall, R(21) = .645, $p < .001$, form recognition and form recall, R(21) = .668, $p < .001$, form recall and meaning recall, R(21) = 0.476, $p < 0.05$. Upon running all correlations a second time, using bootstrapped confidence intervals to account for the small sample size, no differences were found.

Table 4. Correlation between age and FinSL-VT raw scores.

>>> Insert Table 4 here <<<

As shown in Table 4 participants' mean performance scores were different across the four FinSL tasks. These differences generated a hierarchy of difficulty based on which children scored the highest in the meaning recognition task, second highest in the form recognition task, second lowest in the form recall task and the lowest in the meaning recall task. In order to investigate whether the same developmental hierarchy of vocabulary knowledge suggested for ASL and BSL could also be found for FinSL we ran a repeated measures analysis of variance (ANOVA), using a one level design with task (*form recall*, *form recognition*, *meaning recall*, *meaning recognition*) as the within-participant repeated measure. Prior to examining the repeated measures ANOVA results, the assumption of sphericity was tested and Greenhouse-Geisser correction was applied. Results showed a main effect of task, F(1,701) = 268.362, $p < .001$, $\eta_\rho^2 = .921$. This is illustrated in Figure 2.

>>> Insert Figure 2 here <<<

Figure 2. FinSL-VT mean scores in percentages by task. Error bars denote 95% confidence intervals.

A series of paired sample t-tests was carried out to compare performance across the four tasks (alpha level reduced to .008 to compensate for multiple (k=6) comparisons). Performance between all tasks was significantly different (p<.01). Post-hoc tests with Bonferroni corrections were carried out to compare performance across the four tasks (alpha level reduced to .008 to compensate for multiple (k=6) comparisons). Results showed that participants scored higher on meaning recognition than form recognition (p<.000); higher on meaning recognition than form recall (p<.000); higher on meaning recognition than meaning recall (p<.000); higher on form recognition than form recall (p<.002); higher on form recognition than meaning recall (p<.000), and higher on form recall than meaning recall (p<.000). Together these findings demonstrate two things: 1. understanding of different form-meaning mappings improves with age and 2. the four different tasks of the FinSL-VT tap the strength of these mappings to different extents, with meaning recognition being the easiest task, followed by form recognition and form recall, and completed by meaning recall as the hardest task.

Concurrent validity was examined by comparing children's performance on the four tasks in FinSL-VT with their FinSL-RST scores. The mean scores for the FinSL-RST are presented in Table 3. Bivariate correlations between each of the vocabulary tasks and performance on the FinSL-RST showed strong, positive correlations with the meaning recognition task, $R(15) = .658$, $p < .01$, form recognition task, $R(15) = .755$, $p < .01$ and form recall task, $R(15) = .743$, $p < .01$ all of which were statistically significant. No significant correlations were found with the meaning recall task. When carrying out partial correlations between the different tasks in FinSL-VT with FinSL-RST and controlling for age, the following correlations remained significant: FinSL-RST

and form recognition task, R(15) = .666, $p < .01$ and FinSL-RST and form recall task, R(12) = .645, $p < .05$. No significant correlations were found with meaning recognition and meaning recall tasks.

Discussion

This study investigated children's vocabulary knowledge in Finnish Sign Language (FinSL). More specifically, the study explored children's understanding of the mapping between phonological form and meaning in single signs in FinSL, using a multi-layered assessment format developed originally for BSL-VT (Author, XXXX). The present study adapted the web-based BSL-VT for Finnish Sign Language (FinSL) to investigate whether the form-meaning mapping model that it draws on also works for assessing vocabulary knowledge in a less studied signed language. Additionally, this paper described the adaptation process of the test instruments and the psychometric properties of the adapted test and presented findings from a small pilot with 24 deaf and hearing native signers. Finally, the article aims to compare the presented findings to the results from other signed languages, namely BSL, and ASL.

Key findings show that children's performance on all four FinSL tasks was correlated even after partialling out age. These results indicate that all tasks tap the vocabulary knowledge of children. Furthermore, findings showed a hierarchy of the degrees of difficulty in four different form-meaning mapping tasks in FinSL-VT. The two meaning recognition tasks were the easiest for test takers and the recall tasks the hardest. This hierarchy is comparable to results previously reported for BSL and ASL (Author, XXXX; Author, XXXX). Although the ASL-VT version contained a smaller number of test items (80) due to a more limited age range (6–10 years), the findings are similar across all three signed languages. These results are important as they strengthen previous claims by Author (XXXX) that children's knowledge of form-meaning mappings is 'not an all-or-nothing phenomenon but depends on what they need to do with that knowledge' (pp. 1024). As is the case with BSL and ASL, vocabulary acquisition in FinSL proceeds incrementally

with the strength of the mapping between form and meaning increasing over time. Equally important, the findings add strength to the validity of the form-meaning mapping model for use with signed languages.

While participants' understanding of different form-meaning mappings in FinSL bears similarity to the reported findings for ASL and BSL, there was one notable difference: Finnish children's performance scores for the meaning recall task were much lower (mean 37.56% of the maximum score) compared to the scores reported for BSL (mean 52.47% of the maximum scores reported in Author, XXXX) and those for ASL (mean 58.83% of the maximum scores in Author, XXXX). The meaning recall task assesses the depth of vocabulary knowledge and describes how the lexical items in a child's vocabulary are connected to each other in form of synonyms, polysemy and collocations. In both the BSL-VT and ASL-VT version meaning recall was noticed to be the most difficult task form compared to the other three tasks by producing the lowest scores (Author, XXXX; Author, XXXX). During the FinSL-VT test administration, the same procedures were followed as for the BSL-VT and the ASL-VT. However, it should be noticed that in the BSL-VT in the meaning recall task the child is asked to supply only one sign with an associated meaning. This was changed for the ASL-VT version (and adapted for FinSL) where the test taker is asked to supply three signs with associated meaning. Thus, any comparison of this task with the ASL-VT needs to be undertaken with great caution. Results for this task from ASL and FinSL showed that children's scores in ASL-VT were not as low as those in the FinSL-VT. One possible explanation for these differences could be the characteristics of the sample population. The participants in the Finnish study included both deaf and hearing children. All children had been acquiring FinSL as their native language but the language acquisition context was bilingual for all the children and, for many children, even multilingual. Preliminary findings from the meaning recall task carried out in spoken Finnish with the nine hearing children of deaf parents revealed higher scores compared to their performance in sign language (Author, forthcoming). In contrast the

participants used by Author and colleagues for the ASL-VT and/or the BSL-VT included only deaf children. Thus, (part of) the differences in findings across the three signed languages might relate to our participants' bimodal bilingual/multilingual status and the structure of their bilingual or multilingual lexicons. However, this needs to be studied in more depth in the future. Another possible explanation might be cultural differences related to assessment in educational settings, as children in Finland are not used to standardized language testing in general and signed language assessment in particular.

The noticeably lower performance scores of Finnish participants on the meaning recall task may also have affected the results of the internal consistency measures. The average alphas of three tasks, namely meaning recognition, form recognition and form recall met the requirements for the coefficient, whereas the significance value for the meaning recall task was low. This is a somewhat contrary result to the findings reported for the ASL-VT as Author (XXXX) found a lower level of Cronbach's alpha for the meaning recognition task (.55) suggesting that older participants reached the ceiling effect and for that reason the task was the least discriminating task compared to the other three tasks. In the FinSL-VT the meaning recognition, form recognition and form recall tasks all met the adequate level of internal consistency based on Cronbach's alpha even though the participants aged between 11 and 15 got all and nearly all questions correct.

The results in measuring the psychometric properties (reliability and validity) of FinSL-VT showed that the FinSL version of the vocabulary task measures children's vocabulary knowledge in similar ways to the BSL-VT and ASL-VT. Results add new insights into the adaptation process of tests from one signed language to another and show it to be a reliable and valid way to develop assessment tools in lesser researched sign languages such as FinSL.

Conclusions

Standardized assessment tools that are typically developed and normed for hearing monolinguals might reliably show language skills of children sharing similar monolingualism and language acquisition circumstances (Flucher, 2015). However, previous studies on the BSL-VT (Author, XXXX), the ASL-VT (Author, XXXX) and, most recently, the FinSL-VT, show that the underlying format of BSL-VT is a suitable, valid and reliable tool to assess the signed language vocabulary knowledge of children aged 4-15 years and can be used with children acquiring different signed languages (FinSL, BSL or ASL), including children acquiring different levels of bi- and multilingualism and with different levels of hearing. Thus, the reported work on FinSL makes an important contribution to the field of signed language assessment as well as to research on signed language development: it adds to our understanding of signing children's vocabulary knowledge and facilitates cross-linguistic comparisons of this knowledge across different signed languages, including those that are less researched.

Adapting assessment tools from one sign language to another also gives important knowledge about the cross-cultural features between the languages. When adapting the BSL-VT for ASL, considerably fewer items needed to be modified compared to the adaptation for FinSL. One reason for the increased modification is the phonological structure of the signs. In addition to being phonologically related to the target item, the phonological distractor item also needed to be depictable. Currently, the BSL-VT is in the process of being adapted for HongKong Sign Language. This means that there will be three adapted versions of BSL-VT, including signed languages that are very different from each other. The data collected in the process of these adaptations will provide important international research collaboration and important knowledge on cross-linguistics, cross-cultural similarities and differences that can offer an significant insights to theoretical issues on signed language acquisition and linguistical studies in general.

References

Allen, T. E. (2014). The deaf community as a "special linguistic demographic": Diversity rather than disability as a framework for conducting research with individuals who are deaf. In E. Orfanidou, B. Woll & G. Morgan (Eds.), *Research methods in sign language studies: A practical guide* (pp. 21–40). Wiley-Blackwell.

Anderson, D. (2006). Lexical development of deaf children acquiring signed languages. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 135–160). Oxford University Press.

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education, 7*(2)*,* 83–106. https://doi.org/10.1093/deafed/7.2.83

Bizer, S., & Karl, A.-K. (2002a). *Perlesko: Prüfverfahren zur Erfassung lexikalisch-semantischer Kompetenz gehörloser Kinder im Grundschulalter. Manual.* Hamburg: Fachbereich Erziehungswissenschaften, Universität Hamburg.

Bizer, S., & Karl, A.-K. (2002b). *Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schrift- und Lautsprache.* Unpublished doctoral dissertation, Fachbereich Erziehungswissenschaften, Universität Hamburg.

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

Chen Pichler, D., Lillo-Martin, D., & Palmer, J. L. (2018). A Short Introduction to Heritage Signers. *Sign Language Studies*, *18*(3), 309–327.

Clark, E. (2009). *First language acquisition* (2nd ed.). Cambridge University Press.

Dickinson, D.K., & Porche, M.V., (2011). Relationship between language experiences in preschool classroom and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 83*(3), 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. American Guidance Service.

Dunn, L.M., Dunn, L.M., & NFER-Nelson. (1997). *The British Picture Vocabulary Scale Second Edition*. NFER-Nelson, UK.

Enns, C., & Herman, R. (2011). Adapting the assessing British Sign Language Development: Receptive Skills Test into American Sign Language. *Journal of Deaf Studies and Deaf Education, 16*(3), 362–374. https://doi.org/10.1093/deafed/enr004

Fenson, L., Marchman, V. A., Thal, J. D., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Brookes Publishing.

FinSL-VT. (2019). Retrieved January 21, 2020, from https://viittomatesti.cc.jyu.fi/.

Fulcher, G. (2015). *Re-examining language testing. A philosophical and social inquiry.* Routledge.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Walters, J. (2012). *MAIN: Multilingual assessment instrument for narratives* Zentrum für Allgemeine Sprachwissenschaft.

Hadley, E. B., & Dickinson, D. K. (2020). Measuring young children's word knowledge: a conceptual review. *Journal of Early Childhood Literacy*, *20*(2), 223-251. https://doi.org/10.1177/1468798417753713

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

Hermans, D., Knoors, H., & Verhoeven, L. (2010). Assessment of sign language development: The case of deaf children in the Netherlands. *Journal of Deaf Studies and Deaf Education*, *15*(2), 107-119. https://doi.org/10.1093/deafed/enp030

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development, 74*(5), 1368–1378. doi:10.1111/1467-8624.00612

Jantunen, T. (2017a). Fixed and NOT free: Revisiting the order of the main clausal constituents in Finnish Sign Language from a corpus perspective. *SKY Journal of Linguistics, 30,* 137–149.

Jantunen, T. (2017b). Constructed action, the clause and the nature of syntax in Finnish Sign Language. *Open Linguistics, 3*(1), 65–85. doi:10.1515/opli-2017-0004

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorial data. *Biometrics, 33*, 159–174.

Laufer, B. (2013). Vocabulary and writing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21*, 202–226. https://doi.org/10.1191/0265532204lt277oa

Lederberg, A. R., & Spencer, P. E. (2009). Word-learning abilities in deaf and hard-of-hearing preschoolers: Effect of lexicon size and language modality. *Journal of Deaf Studies and Deaf Education*, *14*(1), 44–62. https://doi.org/10.1093/deafed/enn021

Lillo-Martin, D. (1994). Setting the null argument parameters: Evidence from American Sign Language and other languages. In B. Lust, M. Suner, & G. Hermon (Eds.), *Syntactic theory and first language acquisition: Cross linguistic perspectives: Vol. 2*. Anaphora (pp. 301–318). Erlbaum.

Lin, L. H. F., & Morrison, B. (2010). The impact of the medium of instruction in Hong Kong secondary schools on tertiary students' vocabulary. *Journal of English for Academic Purposes, 9*(4), 255–266. http:// 10.1016/j.jeap.2010.09.002

Lu, J., Jones, A., & Morgan, G. (2016). The impact of input quality on early sign development in native and non-native language learners. *Journal of Child Language*, *43,*(3), 537-552. https://doi.org/10.1017/S0305000915000835

Marschark, M. (1997). *Raising and educating a deaf child*. Oxford University Press.

Mayberry, R.I. (1994). The Importance of Childhood to Language Acquisition: Insights from American Sign Language. In: J.C. Goodman & H.C. Nusbaum, (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Words* (pp. 57–90). MIT Press.

Mayberry, R.I. (1993). First-Language Acquisition After Childhood Differs from Second-Language Acquisition: The Case of American Sign Language. *Journal of Speech and Hearing Research*, *36*, 1258–70.

Mayberry, R.I. & Eichen, E.B. (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.

Miralpeix, I. (2019). L1 and L2 vocabulary size and growth. In S. Webb (Ed.) *Routledge Handbooks in Vocabulary Studies* (pp. 189–206). Routledge

Mitchell, R. E., & Karchmer, M. A. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies, 4*(2),138–163.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and word likeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*(5),1033–1047. doi:10.1044/1092-4388(2005/072)

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6–19). Cambridge University Press.

Nunnally, D.J., (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. McGraw-Hill.

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing*, *25*(7), 1635–1664. https://doi.org/10.1007/s11145-011-9351-6

Puupponen, A. (2019). *Understanding nonmanuality–A study on the actions of the head and body in Finnish Sign Language*. JYU dissertations.

Qian, D.D. & Lin, L. H. F. (2019). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.) *Routledge Handbooks in Vocabulary Studies* (pp. 66–80). Routledge.

Reynolds, W. (2018). Young bimodal bilingual development of referent tracking in signed narratives: Further justification of heritage signer status. *Sign Language Studies, 18*(3), 328–354.

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, *21*(5), 617–635. https://doi.org/10.1177/1367006916643528

Rowe, M. L., & Leech, K. A. (2019). A parent intervention with a growth mindset approach improves children's early gesture and vocabulary development. *Developmental science*, *22*, e12792. https://doi.org/10.1111/desc.12792

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212–226. https://doi.org/10.1017/S0261444815000075

Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *53*(1), 146–159. https://doi.org/10.1044/1092-4388(2009/08-0160)

Sheng, L., Pena, E. D., Bedore, L. M., & Fiestas, C. E. (2012). Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language and Hearing Research, 55*(1), 1–15. https://doi.org/10.1044/1092-4388(2011/10-0254)

Singleton, J.L. & Newport, E.L. (2004). When Learners Surpass their Models. The Acquisition of American Sign Language from Insufficient Input. *Cognitive Psychology*, *49*(4), 370–407. https://doi.org/10.1016/j.cogpsych.2004.05.001

Stokoe, W. C. (1960). *Sign language structure*. Silver Spring.

Takkinen, R. (2005). Some observations on the use of HamNoSys (Hamburg Notation System for Sign Languages) in the context of the phonetic transcription of children's signing. *Sign Language & Linguistics*, *8*(1/2), 99–118. https://doi.org/10.1075/sll.8.1.05tak

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong and N. Meir (Eds), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 301–330). Multilingual Matters.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Ibrahim, L.A., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication disorders, 53*, 888–904.

Unsworth, S. (2013). Assessing the role of current and *cumulative* exposure in simultaneous Bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition, 16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Vinson, D., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity and iconicity. *Behavior Research Methods*, *40*, 1079–1087. https://doi.org/10.3758/BRM.40.4.1079

Wauters, L. N., Tellings, A. E., Van Bon, W. H., & Van Haaften, A. W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, *24*, 385–406. https://doi.org/10.1177/0265532215575627

Wauters, L. N., Tellings, A. E., Van Bon, W. H., & Mak, W. M. (2007). Mode of acquisition as a factor in deaf children's reading comprehension. *Journal of Deaf Studies and Deaf Education*, *13*(2), 175-192. https://doi.org/10.1093/deafed/enm050

Woolfe, T., Herman, R., Roy, P., & Woll, B. (2010). Early vocabulary development in deaf native signers: A British Sign Language adaptation of the communicative development inventories. *Journal of Child Psychology and Psychiatry, 51*(3), 322–333. https://doi.org/10.1111/j.1469-7610.2009.02151.x

Yanagisawa, A. & Webb, S. (2019). Measuring depth of vocabulary knowledge. In S. Webb (Ed.) *Routledge Handbooks in Vocabulary Studies* (pp. 371–386). Routledge.

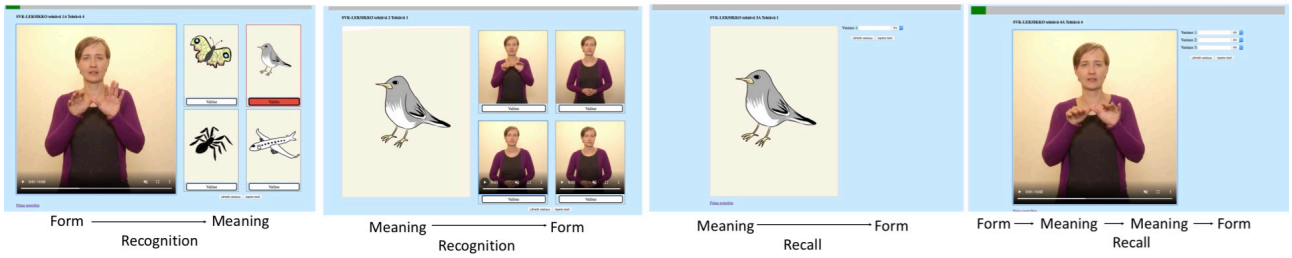

Figure 1. *FinSL-VT tasks.*

Table 1. Stages for adaptation process from BSL_VT to FinSL-VT based on Enns and Herman (2011) and Author (XXXX)

1. Consultation with FinSL panel, consisting of deaf and hearing experts all of whom were fluent or native signers*
2. Development of new test items and distractors*
3. Adjustment of items/images to account for cultural differences*
4. Recording of new target signs and distractors, introduction and instructions. Drawing the target pictures and distractors*
5. Validation of target items**
6. Development of the web-based format**
7. Pre-pilot of FinSL Vocabulary test prototype with two deaf adults*
8. Pilot I with complete set of items*
9. Revision of items based on the pilot data and comments from the panel members

* = the stage that was established in the adaptation process of both Enns and Herman (2011) and Author (XXXX)

** = the stage that was established in the adaptation process of Author (XXXX).

Table 2. Demographic information of participants*.

N	Age			Hearing status of the children	Hearing status of the children's parents	The number of different languages the child uses and exposes daily reported by the parents
	Min-Max	Mean	SD			
24	4;1–15;7	9;11	3.6	8 deaf 16 hearing	20 had both parents deaf 4 had one parent deaf one hearing	11 FinSL and Finnish 2 FinSL and another SP 5 FinSL, Finnish, another SP 2 FinSL, another SL, Finnish 4. FinSL, another SL, Finnish, another SP

* In order to protect the anonymity of the children studied other sign languages than FinSL and other spoken languages than Finnish acquired by the child is only referred 'another sign language (SL)' and another spoken language (SP)'.

Table 3. Raw scores for the FinSL-VT tasks and FinSL-RST task

Task	Mean	SD	Minimum-Maximum
Meaning recognition	104,08 (86.73%)	15.35	67–120
Form recognition	93,45 (77.88%)	18.80	64–112
Form recall	86,83 (72.36%)	15.22	52–113
Meaning recall	135,21 (37.56%)	63.44	17–271
FinSL-RST (n=15)	33,80 (71.91%)	5.91	22–41

Meaning recognition, Form recognition and Form recall tasks have a maximum score of 120, Meaning recall task has a maximum score of 360. FinSL-RST has a maximum score of 47.

Table 4. Correlation between age and FinSL-VT raw scores.

Variable	Meaning recognition	Form recognition	Form recall	Meaning recall
Age	0.80** (.67-.90)	0.77** (.58-.90)	0.78** (.64-.88)	0.76** (.60-.89)

**p<.01, BCa bootstrap 95% CIs reported in parenthesis.

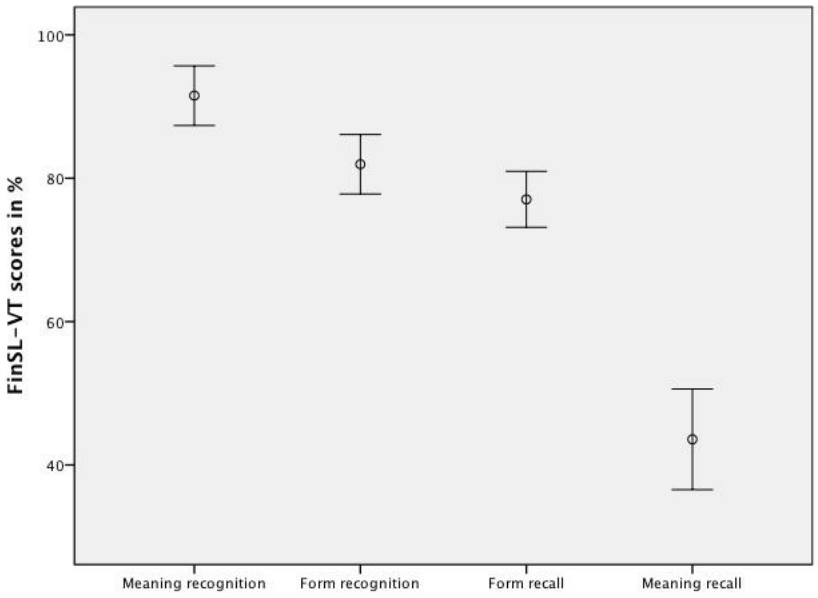

Figure 2. FinSL-VT mean scores in percentages by task. Error bars denote 95% confidence intervals.