Development of a Technology-Assisted Assessment for Sign Language Learning

https://doi.org/10.3991/ijet.v17i06.26959

Tobias Haug[1](✉), Sarah Ebling[2], Katja Tissi[1], Sandra Sidler-Miserez[1], Penny Boyes Braem[3]
[1] University of Teacher Education in Special Needs (HfH), Zurich, Switzerland
[2] Department of Computational Linguistics, University of Zurich, Zurich, Switzerland
[3] Center for Sign Language Research, Basel, Switzerland
tobias.haug@hfh.ch

Abstract—The goal of a recently concluded project in Switzerland was to pioneer an assessment system for lexical signs of Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS) that relies on automatic sign language recognition. The assessment system gives adult L2 learners of DSGS feedback on the correctness of the manual parameters of signing (handshape, hand position, location, and movement) of isolated signs they produce. In its initial version, the system includes automatic feedback for a subset of a DSGS vocabulary size production test consisting of approximately 100 lexical items at CEFR level A1. The paper at hand reports on the process of selecting the items for the test, compiling training data for the SLR system, and linguistically analyzing errors in the resulting video recordings.

Keywords—sign language assessment, Swiss German Sign Language, linguistic error analysis, sign language recognition, vocabulary production test, L2 learning

1 Introduction

The implementation and the use of the Common European Framework of Reference for Languages [1] is a rather new development in the field of learning sign languages as a second or foreign language in tertiary education in Europe. It has only been with recent attempts to align sign language curricula to the CEFR that the development of assessment instruments to evaluate adult learners of a sign language has become possible. Evidence for this are European projects such as *D-Signs* [2] or *ProSign: Sign Language for Professional Purposes* [3].

In Switzerland, three sign languages are used: Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS), French Sign Language of Switzerland (*Langue des Signes Française Suisse*, LSF-S), and Italian Sign Language of Switzerland (*Lingua dei Segni Italiana Svizzera*, LIS-S). While DSGS is taken to be a language of its own, LSF-S and LIS-S are generally seen as varieties of the sign languages used in France and Italy, respectively. DSGS is the primary language for approximately

5,500 deaf sign language users and a second language to approximately 13,000 hearing persons [4]. The group of hearing learners include hearing children of deaf adults, sign language interpreters, teachers for the deaf, and others. DSGS is composed of five dialects that originated in former schools for the deaf. The differences between the dialects are primarily lexical and pertain, e.g., to semantic fields such as food (distinct signs for regional food items, such as specific breads) and date specifications (distinct signs for weekdays and months) [5].

Aligning existing curricula to the CEFR has become an important topic for DSGS, and subsequently, the assessment of adult learners has gained more attention [6] [7]. Due to the visual-spatial modality of sign languages, video technology was always central to sign language learning and assessment. Given the technological developments of the last thirty years, web-based sign language test delivery and scoring have increased but still pose technical challenges such as poor Internet connectivity, lack of technical support at test site, or lack of storage for online video recordings [8]. Sign language tests make use of web-delivered test formats to assess receptive, productive, and interactive skills in a sign language [9]. Studies reporting on web-based test delivery as part of a larger study assessing receptive sign language skills exist [10] [11], but no research addresses issues such as automatic scoring of signed productions to support learning. This article reports on a prototype for automatic scoring of signed productions. More precisely, we introduce a system developed to assess the vocabulary size knowledge of adult learners of DSGS that is based on automatic sign language recognition (SLR). Sign language recognition so far has mostly been applied at the level of isolated signs [12]; hence, in light of the state of the art of this technology, assessment on the supralexical level was not targeted.

As part of this study, we pose the following research question: What were the linguistic, technical and language testing-related challenges in developing a prototype of an automatic assessment system for DSGS? In doing so, we report on the process of selecting the items for our DSGS test, compiling training data for the SLR system, and linguistically analyzing errors in the resulting video recordings.

2 Literature review

2.1 Sign language linguistics

An important feature in any sign language is the distinction between *manual* and *non-manual* components. Manual components are produced with the hands and arms; non-manual components are produced with the face (e.g., with mouth, cheeks, eyes, eyebrows, etc.), the head, and the upper torso [13]. For example, raised eyebrows can be applied to turn a declarative into an interrogative sentence, and eye gaze can be used to re-establish reference in signing space [14]. With the exception of the use of the mouth, few isolated signs have mandatory non-manual components at the lexical level. It is mostly at other linguistic levels, e.g., that of syntax, that non-manual information comes into play.

Hence, for the work presented in this paper, the manual components of signing are of predominant interest. They are typically divided into the subcomponents handshape (the form of the hands, e.g., a fist, flat hand, etc.), hand position (the orientation of the hand), location (where the manual activity is performed), and movement (an optional motion inherent in the sign). These four subcomponents are comparable to phonemes in spoken languages in that they are capable of producing distinctions in meaning [13].

2.2 Assessing and scoring vocabulary knowledge of adult learners

Most publications dealing with sign language tests target deaf children acquiring a sign language as a first language (for an overview, see, e.g., [15] [16]). Among them are a number of publications dealing with the development of vocabulary tests (receptive picture-naming tasks), either as an integrated part of a larger test battery (e.g., for Sign Language of the Netherlands: [17]) or as an independent vocabulary test to evaluate the strength of vocabulary knowledge in deaf children [18] [19]. All of these tests have in common that they use as criterion of correctness [20] a right/wrong distinction. The two vocabulary tests mentioned above use a (offline) computer- and web-based format, respectively.

95% of deaf sign language users are born to hearing parents [21]. This renders learning a sign language as a second language (L2) by hearing adults (among which are parents of deaf children) an important topic. Adult L2 learners of sign are M2L2 (second-modality second-language) learners in contrast to M1L2 (first-modality second-language) learners of sign, who are deaf learners of a sign language other than the sign language they first acquired [22]. As [23] state, "scant research has been carried out on either the psychological processes of acquiring a signed language as a hearing adult or the efficacy of particular teaching methods or approaches…This means that teachers often lack an evidence base from which to make decisions about how to go about teaching sign languages to adult learners" (p. 323).

With regard to assessment, only few publications introduce test instruments for adult learners of a sign language. One example of such an instrument is the Sign Language Proficiency Interview (SLPI) for American Sign Language (ASL) [24], which is an adaptation of the Oral Proficiency Interview for English. The scoring instrument of the SLPI includes the criterion of "vocabulary knowledge" but only as a very broadly defined construct across different levels. Other examples of sign language tests for adult learners are the Sentence Reproduction Test for ASL [25] as a "global, objective assessment ASL proficiency test" (p. 171) or the ASL Discrimination Test [10], which tests "learners' ability to discriminate phonological and morphophonological contrasts in ASL, [and] provides an objective overall measure of ASL proficiency" (p. 473). These tests target a variety of constructs (communicative competence, global measure of ASL, (morpho-)phonological contrasts) but are not measuring the construct of vocabulary.

2.3 Sign language recognition for sign language learning and assessment

In the wake of digital transformation, sign language learning and instruction have undergone serious changes. For example, due to the lack of immediately available conversation partners with a high command of a given sign language, many learners now turn to the Internet to consult sign language dictionaries and other resources (e.g., fingerspelling tutors) or obtain videos of sign language utterances [26]. While these are examples of the usage of general technology (e.g., videos) for computer-assisted language learning (CALL) [27] [28], automatic sign language recognition (SLR), the identification of the form and meaning of isolated signs or sign sequences, in particular, can be useful for sign language learning and assessment.

Early systems, administered to both children [29] [30] [31] and adults [32] [33] [34] [35] suffered from the comparatively low state of the art of SLR research. Later systems focused on specific aspects of system development. For example, [36], [37], and [23] envisioned a system, entitled *My Interactive Auslan Coach* (MIAC), that provides automatic feedback on the correctness of the handshape and movement of Australian Sign Language (Auslan) signs produced by hearing adult learners of the language. Feedback was intended to be based on automatic SLR via a Kinect sensor. However, apart from a prototype, the system does not appear to be operational; instead, the focus of the project seems to have been on questions pertaining to user-centered design (UCD) [36].

[38] proposed a system that automatically analyzes the productions of adult learners of American Sign Language (ASL) and provides immediate feedback on both manual and non-manual components of signing (on the isolated-sign or sentence level). An important contribution of this research lies in exploring the optimal design of the feedback. By conducting controlled experiments in a Wizard-of-Oz setting,[1] in which humans mimicked the functionality of an automatic SLR and assessment component and manually produced standardized feedback messages, the authors found that learners preferred time-synchronized feedback (presented in the form of time-aligned annotations of a video of the learner's signing performance with summary notes at the end) over non-synchronized feedback (summary notes at the end of video only). Learners' performances, measured after presenting the learners with the feedback and asking them to repeat the signing performance, were also higher when they were shown a video with feedback (time-synchronized with summary notes at the end or summary notes at the end only) than without feedback. In a post study, learners emphasized the benefit of enhancing videos of their performances with photos of correct aspects of a sign (e.g., a targeted facial expression or a movement).

3 Development of a vocabulary production test for adult learners of DSGS

While SLR has been applied to sign language learning, the combination with sign language assessment is new. The assessment system that we present here provides adult

[1] *Wizard of Oz* experiments more generally refer to experiments in which a human wizard takes over the role of the machine while leading the subjects to believe that he/she is a machine.

L2 learners of DSGS with feedback on the correctness of the manual parameters (handshape, hand position, location, movement) of isolated signs they produce. In its initial version, the system includes automatic feedback for a subset of a DSGS vocabulary size production test for CEFR level A1 consisting of approximately 100 lexical items (see Appendix A). As part of the testing scenario, learners are prompted with a DSGS gloss (sign language glosses are capitalized spoken language words used as labels for semantic aspects of signs, e.g., HOUSE or CAR) on a monitor in front of them. They then produce the sign while their production is recorded by a video camera in front of them. Following this, they receive feedback from the automatic assessment system.

3.1 Selection of test items

Only signs of the native, conventional lexicon were considered for the vocabulary test. In order to arrive at a concept comparable to that of *word families* [39], one could include signs that involve morphological changes to the lexical base form [40]. However, the problem remains that this group of signs is less clearly defined for sign languages than for spoken languages, which would have an impact on the definition of what is a correctly produced sign in a DSGS vocabulary test. Just considering sign types that are known to have a stable form-meaning relationship [41] is further complicated by the fact that there exists little research on acceptable phonetic variations of signs (for an exception, see [27]).

In absence of a sufficiently large corpus of continuous signing of DSGS, it was not possible to select the items for the vocabulary production test based on the criterion of corpus frequency, such as is typically done in the case of spoken languages (for English, e.g., [42]). Item selection was therefore based on existing DSGS teaching materials [43], [44], [45], and [46] known to correspond to CEFR level A1. The DSGS teaching materials are used as part of four levels of DSGS courses offered by the Swiss Federation of the Deaf, each course consisting of 30 lessons (total: 120), which provides a rough estimation for the CEFR level A1. The number of sign types available in the DSGS teaching materials is approximately 3,800 [47]. Sign types are defined in our study as signs with a stable form-meaning relationship [41], that do not include morphological modifications (and test items are the sign types used in a vocabulary test). In order to reduce this number to approximately 100 (test and practice items), the following steps were applied [48]:

1. Removal of name signs, i.e., signs for persons (e.g., CHARLY CHAPLIN), organizations (e.g., name of a university), and places (e.g., country names), as many of these are borrowed from other sign languages.
2. Removal of body-part signs like NASE (‘nose’), as these are often produced by merely pointing at the respective body part, i.e., using an *indexing technique*.
3. Removal of pronouns like DU (‘you’), as they also correspond to indexical signs.
4. Removal of number signs, as they tend to have several regional variants, e.g., the number sign ELF (‘eleven’).
5. Removal of signs making use of fingerspelling, like the sign JANUAR (‘January’), which involves the letter J from the DSGS manual alphabet.

6. Removal of signs composed of multiple successive elements, as most of these signs also occurred in the DSGS teaching materials as separate lexemes. For example, the sign ABENDESSEN ('dinner') is composed of the two signs ABEND ('evening') and ESSEN ('meal'), both of which are also contained in the list of sign types of the DSGS teaching materials.
7. Removal of old signs, as current DSGS learners cannot be expected to know them.
8. Removal of productive forms. The reason for this step was that the phonological parameters of productive signs tend to be variable, which poses an undue challenge to the sign recognition system that is part of the assessment framework discussed in this article.
9. Removal of signs appearing in fewer than four of the five DSGS dialects.
10. Reduction of manual homonymy: Since the goal was to have as many different sign forms in the vocabulary test as possible, form-identical signs were identified (e.g., BRUDER ('brother'), SCHWESTER ('sister'), and GLEICH ('same')) and only one chosen for the test. If applicable, preference was given to that sign which was contained in a list of 1,000 common sign concepts (Efthimiou et al., 2009).
11. Removal of signs that are very similar to well-known co-speech gestures, such as the sign SUPER, which corresponds to a thumb-up gesture.
12. Removal of signs with German glosses that are lexically ambiguous. For example, the German word AUFNEHMEN can have the meaning of *record* or *accept/include*, concepts which in DSGS are expressed with two separate signs. In cases like these, test takers confronted with the German gloss AUFNEHMEN would not know which sign to produce.
13. Prioritization of concepts that also occurred in studies investigating familiarity or subjective frequency ratings for BSL [49] and ASL [50] [51] and in a list of 1,000 sign concepts [52].

In this way, the 3,800 sign types from the DSGS teaching materials were reduced to a set of approximately 100 test items. The item set was not balanced with respect to parts of speech, as is often done when sampling items for a spoken language vocabulary test. This was because the question of whether the concept of parts of speech can be applied to sign languages is still an unsolved one within sign linguistics [53]. Table 1 summarizes the item selection process.

Table 1. Summary of the item selection process

Removed
Name signs: persons, organizations, places, languages
Body-part signs
Pronouns
Number signs
Primarily fingerspelled components
Signs composed of multiple successive segments
Old signs
Productive signs
Signs appearing in less than four of the five DSGS dialects
Homonyms
Signs overlapping with co-speech gestures

Signs with ambiguous German glosses Signs with occurrence <3 in DSGS corpora
Prioritized
Signs with concepts in [52] Signs for concepts included in all of the following studies: [49], [50], and [51]

3.2 Compilation of training data

State-of-the-art approaches to SLR are based on deep learning methods that require large amounts of data. To provide the SLR component of the automatic DSGS assessment system with sufficient samples to learn from ("training data"), a dataset containing videotaped repeated productions of the 100 items of the DSGS vocabulary test with associated transcriptions and annotations was created, consisting of data from 11 adult L1 signers and 19 adult L2 learners of DSGS.

Recording procedure. The signing performances were recorded with six different visual sensors in a studio environment: a Microsoft Kinect sensor to obtain skeleton and depth information, two GoPros (one with a high framerate to capture fast movements in signing and the other with a high resolution to capture details in the face of the signer), and three HD cameras capturing different perspectives (top, left, right; the front perspective was taken from the color image of the Kinect; see Figure 1).

Fig. 1. Set-up of the studio

The focus of the data collection process described in this article was on obtaining training data for the SLR system. Therefore, in an attempt to reduce the number of instances in which no sign was produced at all, the participating signers were provided with the test items prior to the recordings in the form of a list of glosses with accompanying German example sentences. Table 2 shows a selection of glosses along with context examples. The sentences had been gathered from a DSGS online lexicon[2] and, where necessary, shortened and modified. The rationale behind providing German example sentences in addition to DSGS glosses was to further reduce any semantic ambiguity remaining even after clearly ambiguous glosses had been eliminated in the item selection process.

Table 2. Glosses and example sentences

Gloss	Example sentence
ANGESTELLT ('EMPLOYED')	Sie ist in einer grossen Firma angestellt. ('She is employed by a large corporation.')
THEATER ('THEATRE')	Das Theater findet in Basel statt. ('The theatre play takes place in Basel.')
WARTEN ('WAIT')	Ich warte, bis der Arzt kommt. ('I am waiting for the doctor to come.')

Upon recording, participants were asked to perform each of the 100 signs three times. The glosses with German example sentences served as prompts for the first two passes, while the prompt for the third pass was a video of a signer performing the sign. The video corresponded to the base form of the sign in a DSGS lexicon [47]. Participants were asked to mirror the sign they saw in the video, not repeat a potential dialect variant that they might have produced in the previous two passes. The order of the signs in the three passes was different and participants were asked to return to a neutral position after each sign. They were not required to look into a particular camera but rather direct their eye gaze towards the general area of the cameras. Participants were specifically instructed to sign the base forms of the lexical items, not modified versions based on the context evoked in the example sentences. Recordings lasted between 30 and 45 minutes.

While the DSGS vocabulary production test is ultimately aimed for use by L2 learners, the goal of the recordings described here was to obtain both L1 and L2 data for training the recognition and assessment system. 11 L1 and 19 L2 signers participated in the recordings. The L1 participants were recruited by the deaf members of the project team; they were native DSGS signers and/or trained DSGS instructors. To recruit L2 participants, a call for participation was released via various channels. L2 participants had to have completed four courses in the course framework of the Swiss Federation of the deaf corresponding to parts of CEFR level A1. Both L1 and L2 participants were asked to complete a background questionnaire prior to the recordings. The background questionnaire was a modified version of a questionnaire developed in the DGS Corpus Project [55].

[2] https://signsuisse.sgb-fss.ch/ (last accessed September 7, 2017)

Linguistic annotation. To perform transcription and annotation on the videos obtained through the procedure previously outlined, the videos were postprocessed and imported into iLex, a software for creating and analyzing sign language lexicons and corpora [55]. In iLex, all occurrences of a sign in a transcript (sign tokens) are linked back to their sign type in the lexicon, and changes of the sign type affect all sign tokens in all transcripts.

3.3 Criterion of correctness for the assessment system

Members of the project manually annotated information for the second pass. If a sign was produced multiple times in this pass (recall that self-correction was permitted during the recordings), only the last attempt was considered. A two-person principle was observed, i.e., each annotation produced by one annotator was checked by another. Each production of an individual sign was classified into one of six categories:

1. Same lexeme as target sign: same meaning, same form
2. Same lexeme as target sign: same meaning, slightly different form
3. Same lexeme as target sign: same meaning, different form
4. Same lexeme as target sign: slightly different meaning, slightly different form
5. Different lexeme than target sign: same meaning, different form
6. Different lexeme than target sign: different meaning, different form

Instances of Category 1 were sign productions that are identical to the target sign, i.e., to the base form as produced in the model video. Sign productions assigned to Category 2 had the same meaning as the target sign and a slightly different but acceptable form.[3] For example, the sign SPRACHE (‘LANGUAGE’) might have been produced in a slightly different location. Members of Category 3 were judged by the annotators to differ clearly and significantly from acceptable variant forms (cf. below for the link between categories and test decisions, i.e., decisions regarding the correctness of the productions). For example, if SPRACHE, which has an open handshape, was produced with a closed handshape, this occurrence was labeled with Category 3. Instances of Category 4 were morphophonemic/semantic variants, e.g., modifying SPRACHE from singular to plural, resulting in a slightly different form and slightly different meaning. Sign productions that represented dialect variants were assigned to Category 5, indicating identical meanings but different forms. Sign productions with both an entirely different meaning and form, e.g., productions of the sign BAUM (‘TREE’) for the prompt SPRACHE, were assigned to Category 6.

Table 3 shows the mapping of category assignments to test decisions: Members of Categories 1, 2, 4, and 5 are rated as correct, while members of Categories 3 and 6 are considered incorrect. This information was used to train the assessment system.

[3] These instances are sometimes called *allophonic variants*.

Table 3. Link between category assignments and test decisions

Category	Same lexeme as target sign?	Same meaning?	Same form?	Test decision
1	Yes	Yes	Yes	Correct
2	Yes	Yes	Slightly different	Correct
3	Yes	Yes	No	Incorrect
4	Yes	Slightly different	Slightly different	Correct
5	No	Yes	No	Correct
6	No	No	No	Incorrect

3.4 Linguistic error analysis

Alongside the creation of the training data for the SLR system, the errors committed by the L2 learners of the study (i.e., Category 3 above) were linguistically analyzed to inform future sign language instruction practice. As part of a statistical analysis of single-parameter errors, movement was found to be the parameter most susceptible to errors, followed by location, orientation, and handshape. This was identical to the error hierarchy observed for American Sign Language [57]. The study also conducted an analysis of production errors with respect to combinations of manual parameters, something that previously has not been undertaken. The parameter combination most frequently involved in errors was movement combined with location.

4 Discussion and conclusions

In this paper, we have reported on the development of the first assessment and feedback system for DSGS. The system is based on a vocabulary size production test for CEFR level A1 and provides feedback on a subset of the 100 lexical items of the test to adult learners of DSGS. The system relies on automatic sign language recognition and targets the correctness of the manual parameters of signing (handshape, hand position, location, and movement).

We have discussed how we arrived at a set of 100 items for the test: The items were sampled from existing DSGS teaching materials that correspond to CEFR level A1, by applying different linguistic criteria, such as removing second-person-singular pronouns, old signs, or signs that bear high resemblance to well-known co-speech gestures.

We have described how data to train the sign language recognition component of the system was obtained: Repeated productions of the 100 items of the DSGS vocabulary test were recorded with six different visual sensors in a studio environment. To perform transcription and annotation, the resulting videos were postprocessed and imported into a software for creating and analyzing sign language lexicons and corpora. Each production of an individual sign was classified into one of six categories. Productions by the L2 learners corresponding to categories deemed incorrect were then further linguistically analyzed to inform future sign language instruction practice.

In a next step, the assessment system will be ported to an online system so that learners can assess their DSGS vocabulary knowledge at home and receive immediate feedback.

There are concrete plans in a follow-up study that the system as outlined in this contribution will be extended such that sentence-level assessment becomes possible. This requires additionally considering the non-manual features of signing (such as positions or movements of the head and shoulder, eyebrow, mouth, eyes, eye gaze, etc.), which are particularly pertinent for marking sentence types, in the recognition process.

The mid- and long-term goal is to use this technology for both formative and summative assessment purposes. For example, to have an assessment scenario where a learner can check his or her vocabulary knowledge with a self-assessment system remotely by signing into a webcam, the signal is sent to a server where the system provides immediate feedback on the produced sign. In a summative testing scenario, the learners can take a sentence-level exam (e.g., part of a module exam) where he or she also receives automatic feedback on their signed production.

5 Acknowledgment

This work was funded by the Swiss National Science Foundation Sinergia project "Scalable Multimodal Sign Language Technology for Sign Language Learning and Assessment (SMILE)", grant agreement number CRSII2 160811.

6 References

[1] Council of Europe, Common European Framework of Reference for languages: Learning, teaching, assessment. Companion volume. Council of Europe Publishing, 2009.

[2] L. Leeson and C. Grehan, Presentation, Topic: "A Common European Framework for sign language curricula? D-sign(ing) a curriculum aligned to the Common European Framework of Reference". D-Signs Project, Bristol, UK, 2009.

[3] L. Leeson, T. Haug, C. Rathmann, B. Van den Bogaerde, and N. Sheneman, The implementation of the Common European Framework of Reference (CEFR) for signed languages in higher education – Results of an international survey [Survey report], 2018.

[4] P. Boyes Braem, Class Lecture, Topic: "Overview of research on the signed languages of the deaf", Oct. 8, 2012.

[5] P. Boyes Braem, "Studying Swiss German Sign Language dialects", in *Recent research on European sign,* F. Loncke, P. Boyes Braem, and Y. Lebrun, Eds. Swets and Zeitlinger, 1984, pp. 93-103.

[6] T. Haug and S. Ebling, "Using open-source software for sign language learning and assessment: The case of a web-delivered Yes/No vocabulary test for Swiss German Sign Language", *International Journal of Emerging Technologies in Learning (IJET)*, vol 14, no. 19, pp. 188-196, 2019. https://doi.org/10.3991/ijet.v14i19.11123

[7] T. Haug, S. Ebling, P. Boyes Braem, K. Tissi, and S. Sidler-Miserez, "Sign language learning and assessment in German Switzerland: Exploring the potential of vocabulary size tests for Swiss German Sign Language", *Language Education & Assessment*, vol. *2*, no. 1, pp. 20-40, 2019. https://doi.org/10.29140/lea.v2n1.85

[8] T. Haug, "Use of information and communication technologies in sign language test development: Results of an international survey", *Deafness & Education International*, vol. 17, no. 1, pp. 33-48, 2015. https://doi.org/10.1179/1557069X14Y.0000000041

[9] W. Mann and T. Haug, "Benefits and shortcomings of new technologies for sign language assessment", poster presented at Sign language acquisition and assessment conference, Haifa, Israel, 2018.

[10] J. H. Bochner, V. J. Samar, P. C. Hauser, W. M. Garrison, J. M. Searls, and C. A. Sanders, C. A., "Validity of the American Sign Language Discrimination Test". *Language Testing*, vol. 33, no. 4, pp. 473-495, 2016. https://doi.org/10.1177/0265532215590849

[11] P. C. Hauser, R. Paludneviciene, W. Riddle, K. B. Kurz, K. Emmorey, and J Contreras, "American Sign Language Comprehension Test: A tool for sign language researchers", *Journal of Deaf Studies and Deaf Education*, vol. 21, no. 1, pp. 64-69, 2015. https://doi.org/10.1093/deafed/env051

[12] S. Ebling, N. Camgöz, and R. Bowden, "Use of new technologies in L2 signed language assessment", in *Handbook on language assessment across modalities*, T. Haug, W. Mann, and U. Knoch, Eds. New York: Oxford University Press, forthcoming.

[13] P. Boyes Braem, *Einführung in die Gebärdensprache und ihre Erforschung*. Seedorf: Signum Verlag, 1995.

[14] R. Pfau and J. Quer, "Nonmanuals: Their grammatical and prosodic roles", in *Sign languages*, D. Brentari, Ed. Cambridge: Cambridge University Press, 2010, pp. 381-403. https://doi.org/10.1017/CBO9780511712203.018

[15] C. Enns, T. Haug, R. Herman, R. Hoffmeister, W. Mann, and L. Mcquarrie, "Exploring signed language assessment tools in Europe and North America", in *Diversity in deaf education*, M. Marschark, V. Lampropoulou, and E. K. Skordilis, Eds. Oxford: Oxford University Press, 2016, pp. 171-218. https://doi.org/10.1093/acprof:oso/9780190493073.003.0007

[16] T. Haug, "Review of sign language assessment instruments", *Sign Language & Linguistics*, vol. 8, no. 1/2, pp. 61-98, 2005. https://doi.org/10.1075/sll.8.1.04hau

[17] D. Hermans, H. Knoors, and L. Verhoeven, "Assessment of sign language development: The case of deaf children in the Netherlands", *Journal of Deaf Studies and Deaf Education*, vol. 15, no. 2, pp. 107-119, 2010. https://doi.org/10.1093/deafed/enp030

[18] W. Mann, P. Roy, and G. Morgan, "Adaptation of a vocabulary test from British Sign Language to American Sign Language", *Language Testing*, vol. 33, no. 1, pp. 3-22, 2015. https://doi.org/10.1177/0265532215575627

[19] W. Mann and C. Marshall, "Investigating deaf children's vocabulary knowledge in British Sign Language: Vocabulary development in sign language", *Language Learning*, vol. *62, no.* 4, pp. 1024-1051. https://doi.org/10.1111/j.1467-9922.2011.00670.x

[20] L. F. Bachman and A. Palmer, *Language testing in practice*. Oxford: Oxford University Press, 1996.

[21] R. E. Mitchell and M. A. Karchmer (2004), "Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States", *Sign Language Studies*, vol. 4, no. 2, pp. 138-163, 2004. https://doi.org/10.1353/sls.2004.0005

[22] D. Chen Pichler, "Sign language development", in *Language development over the lifespan,* K. de Bot and R. Schrauf, Eds. Routledge, 2009, pp. 217-241.

[23] L. Willoughby, S. Linder, K. Ellis and J. Fisher, "Errors and feedback in the beginner Auslan classroom", *Sign Language Studies*, vol. 15, no. 3, pp. 322-347, 2015. https://doi.org/10.1353/sls.2015.0009

[24] W. Newell, F. Caccamise, K. Boardman, and Ray Holcomb, "Adaptation of the Language Proficiency Interview (LPI) for assessing sign communicative competence", *Sign Language Studies*, vol. 41, pp. 311-347, 1983. https://doi.org/10.1353/sls.1983.0012

[25] P. Hauser, T. Supalla, and D. Bavelier (2008), "American Sign Language sentence reproduction test: Development and implications", in *Sign Languages: Spinning and unraveling the past, present and future. TISLR9, forty-five papers and three posters from the 9th. Theoretical Issues in Sign Language Research Conference*, R. Müller de Quadros, Ed. Petrópolis/RJ. Brazil, 2008, pp. 160-172.

[26] L. Willoughby and C. Sell, "Studying a sign language: What are hearing adults doing outside of class?", *Sign Language Studies*, vol. 19, no. 3, 453-478, 2019. https://doi.org/10.1353/SLS.2019.0004

[27] C. Guetl, V. Chang, A. Edwards, and S. Boruta, "Flexible and affordable foreign language learning environment based on Web 2.0 technologies", *International Journal of Emerging Technologies in Learning (IJET)*, vol. 8, no. 2, pp. 16-28, 2013. https://doi.org/10.3991/ijet.v8i2.2488

[28] V. Persson and J. Nouri, "A systematic review of second language learning with mobile technologies", *International Journal of Emerging Technologies in Learning (IJET)*, vol. 13, no. 2, pp. 188-210, 2018. https://doi.org/10.3991/ijet.v13i02.8094

[29] J. Arendsen, *Seeing signs: On the appearance of manual movements in gestures*. Technische Universiteit Delft, 2009.

[30] H. Brashear, V. Henderson, K. Park, H. Hamilton, S. Lee, and T. Starner, "American Sign Language recognition in game development for deaf children". In Eighth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS, 2006. https://doi.org/10.1145/1168987.1169002

[31] H. Brashear, V. Henderson, K. Park, H. Hamilton, S. Lee, and T. Starner, "American Sign Language recognition in game development for deaf children". In Eighth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS, 2006. https://doi.org/10.1145/1168987.1169002

[32] O. Aran, I. Ari, L. Akarun, B. Sankur, A. Benoit, A. Caplier, P. Campr, A. H. Carrilo, A. H., and F. Fanard, "SignTutor: An interactive system for sign language tutoring", *IEEE Multi Media*, vol. 16, no. 1, pp. 81-93, 2009. https://doi.org/10.1109/MMUL.2009.17

[33] J. Zieren, "Visuelle Erkennung von Handposituren für einen interaktiven Gebärdensprachtutor", PhD dissertation, RWTH Aachen, 2007.

[34] J. Zieren and K.-F. Kraiss, "Robust person-independent visual sign language recognition", *Lecture Notes in Computer Science*, vol. 3522, pp. 520-528, 2005. https://doi.org/10.1007/11492429_63

[35] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American Sign Language recognition with the Kinect", in Proc. 13th International Conference on Multimodal Interfaces - ICMI '11, 2011, pp. 279-286. https://doi.org/10.1145/2070481.2070532

[36] J. Fisher, K. Ellis, L. Willoughby, and J. Barca, "Taking a user centred design approach for designing a system to teach sign language", in Proc. 25th Australasian Conference on Information Systems (ACIS 2014), 2014, pp. 1-10.

[37] K. Ellis, J. Fisher, L. Willoughby, and J. Barca, "A design science exploration of a visual-spatial learning system with feedback", in Proc. 26th Australasian Conference on Information Systems (ACIS 2015*)*, 2015, pp. 1-13.

[38] M. Huenerfauth, E. Gale, B. Penly, S. Pillutla, M. Willard, and D. Hariharan, "Evaluation of language feedback methods for student videos of American Sign Language", *ACM Transactions on Accessible Computing*, vol. 10, no. 1, pp. 1-30, 2017. https://doi.org/10.1145/3046788

[39] J. A. S. Read, *Assessing vocabulary*. Cambridge: Cambridge University Press, 2000. https://doi.org/10.1017/CBO9780511732942
[40] T. Johnston and A. Schembri, *Australian Sign Language: An introduction to sign language linguistics*. Cambridge: Cambridge University Press, 2007. https://doi.org/10.1017/CBO9780511607479
[41] S. König, R. Konrad, and G.Langer (2012), "Lexikon: Der Wortschatz der DGS", in *Handbuch Deutsche Gebärdensprache: Sprachwissenschaftliche und anwendungsbezogene Perspektiven*, H. Eichmann, M. Hansen, and J. Heßmann, Eds. Seedorf: Signum, 2012, pp. 111-164).
[42] B. Laufer, C. Elder, K. Hill, and P. Congdon, "Size and strength: Do we need both to measure vocabulary knowledge?", *Language Testing*, vol. 21, no. 2, pp. 202-226, 2004. https://doi.org/10.1191/0265532204lt277oa
[43] P. Boyes Braem, *Gebärdensprachkurs Deutschschweiz, Stufe 1. Linguistischer Kommentar*. Zurich: GS-Media/Schweizerischer Gehörlosenbund SGB, 2004a.
[44] P. Boyes Braem, *Gebärdensprachkurs Deutschschweiz, Stufe 2. Linguistischer Kommentar*. Zurich: GS-Media/Schweizerischer Gehörlosenbund SGB, 2004b.
[45] P. Boyes Braem, *Gebärdensprachkurs Deutschschweiz, Stufe 3. Linguistischer Kommentar.* Zurich: GS-Media/Schweizerischer Gehörlosenbund SGB, 2005a.
[46] P. Boyes Braem, *Gebärdensprachkurs Deutschschweiz, Stufe 4. Linguistischer Kommentar*. Zurich: GS-Media/Schweizerischer Gehörlosenbund SGB, 2005b.
[47] P. Boyes Braem, "A multimedia bilingual database for the lexicon of Swiss German Sign Language", *Sign Language & Linguistics*, vol. 4, no. 1/2, pp. 133-143, 2001. https://doi.org/10.1075/sll.4.12.10boy
[48] S. Ebling, N. C. Camgöz, P. Boyes Braem, K. Tissi, S. Sidler-Miserez, S. Stoll, S. Hatfield, T. Haug, R. Bowden, S. Tornay, M. Razavi, and M. Magimai-Doss (2018), "SMILE Swiss German Sign Language data set", in Proc. 11th Language Resources and Evaluation Conference (LREC 2018), 2018, pp. 4221-4229.
[49] D. P. Vinson, K. Cormier, T. Denmark, A. Schembri, G. Vigliocco (2008), "The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity", *Behavior Research Methods*, vol. 40, no. 4, pp. 1079-1087, 2008. https://doi.org/10.3758/BRM.40.4.1079
[50] N. K. Caselli, Z. S. Sehyr, A. M. Cohen-Goldberg, and K. Emmorey, "ASL-LEX: A lexical database of American Sign Language", *Behavior Research Methods*, vol. 49, no. 2, pp. 784-801, 2017. https://doi.org/10.3758/s13428-016-0742-0
[51] R. I. Mayberry, M. L. Hall, and M. Zvaigzne, "Subjective frequency ratings for 432 ASL signs", *Behavior Research Methods*, 2013. https://doi.org/10.3758/s13428-013-0370-x
[52] E. Efthimiou, S. E. Fotinea, C. Vogler, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and J. Segouat, "Sign language recognition, generation, and modelling: A research effort with applications in deaf communication", in *Universal access in human-computer interaction. Addressing diversity, lecture notes in computer science*, C. Stephanidis, Ed. Heidelberg: Springer, 2009, vol. 5614, pp. 21-30. https://doi.org/10.1007/978-3-642-02707-9_3
[53] S. Erlenkamp, "Lexikalische Klassen und syntaktische Kategorien in der Deutschen Gebärdensprache: Warum das Vorhandensein von Verben nicht unbedingt Nomen erfordert", in *Gebärdensprachlinguistik 2000 – Theorie und Anwendung: Vorträge vom Symposium Gebärdensprachforschung im Deutschsprachigen Raum, Frankfurt a.M., 11.-13. Juni 1999,* H. Leuninger and K. Wempe, Eds. Seedorf: Signum Verlag 2001, pp.67-91.
[54] T. Haug, "Development and evaluation of two vocabulary tests for Swiss German Sign Language", Master thesis, Lancaster University, 2017. https://doi.org/10.13140/RG.2.2.25397.17129

[55] T. Hanke, „Wörterbuch ohne Wörter? Zum Entstehen eines Wörterbuches der Deutschen Gebärdensprache“, in *Jahrbuch der Heidelberger Akademie der Wissenschaften für 2016,* Heidelberger Akademie der Wissenschaften, Ed. Heidelberg: Universitätsverlag Winter, 2017, pp. 84-88. http://www.haw.uni-heidelberg.de/md/haw/publikationen/jahrbuch_2016_www.pdf

[56] T. Hanke and J. Storz, “iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography”, in Proc. LREC 2008 Workshop Proceedings, 2006, pp. 64-67.

[57] D. Schlehofer and I. J. Tyler, “Errors in Second Language Learners’ Production of Phonological Contrasts in American Sign Language”, *International Journal of Language and Linguistics*, vol. 3, no. 2, pp. 30-38, 2016. http://ijllnet.com/journals/Vol_3_No_2_June_2016/4.pdf

7 Authors

Tobias Haug studied sign linguistics at Hamburg University and deaf education at Boston University, where he received his Master in Education in 1998. In 2009 he earned his PhD at Hamburg University. From 1998 to 2004 he worked as a sign language interpreter and researcher. Since 2004 he has been the program director of the sign language interpreter program in Zurich, Switzerland. One of his research interests is sign language assessment in the context of new technologies. In 2017 he received his MA in Language Testing from Lancaster University.

Sarah Ebling is a senior researcher at the University of Zurich. With a background in computational linguistics, her focus in research and teaching is on the contribution of language technology to accessibility for persons with disabilities. Her previous research includes sign language technology, specifically, machine translation into sign language and sign language synthesis, the generation of signing avatars. More recently, she has been involved in an interdisciplinary project centering around automatic sign language recognition.

Katja Tissi, is a senior lecturer in the sign language interpreter training program at the HfH (with the focus on teaching Swiss German Sign Language). She also has been involved in numerous research projects on Swiss German Sign Language over the last 25 years. She is currently involved in a research project on automatic sign language recognition and its application for sign language assessment.

Sandra Sidler-Miserez is a trained sign language teacher and has been involved in different sign language assessment-related research project at the HfH. She is currently involved in a research project on automatic sign language recognition and its application for sign language assessment.

Penny Boyes Braem is a trained linguist and founder of the Center for Sign Language Research in Basel, Switzerland. She is the pioneer of research in Swiss German Sign Language, starting in the 1980ies. Her research focused on sign language acquisition, sign language resources, and the structure of DSGS.

Article submitted 2021-09-17. Resubmitted 2022-01-06. Final acceptance 2022-01-20. Final version published as submitted by the authors.

8 Appendix

Appendix A: Items of DSGS translation test (including test and practice items)

No.	Item (DSGS gloss)	Items (translation)	Sentences
P1	WIDERSPRUCH	'contradiction'	There is a contradiction between your and my opinion.
P2	WIE VIEL	'how much'	How much money do you have?
P3	WOHNEN	'to live'	I live on a farm.
P4	WORT	'word'	I don't understand this word.
P5	ZEIGEN	'to show'	I show you my new computer.
I1	ENTTÄUSCHT	'disappointed'	I'm disappointed that I didn't pass the exam.
I2	ABSCHALTEN	'to relax'	I can relax well while I'm on vacation.
I3	AUSTAUSCHEN	'to exchange'	During our conversation, we had time to exchange ideas.
I4	FUSSBALL	'soccer'	The family watches soccer together.
I5	SUCHEN	'to look for'	I am looking for my glasses.
I6	STRASSE	'street'	This street is new.
I7	SCHÜTZEN	'to protect'	Parents want to protect their children.
I8	BEGLEITEN	'to follow'	I follow you.
I9	VIOLETT	'violet'	I like the violet bag.
I10	ERZÄHLEN	'to tell'	The woman likes to tell fairy tails.
I11	MIT	'with'	I drink coffee with sugar and cream.
I12	PRÜFUNG	'exam'	The girl is nervous before every exam.
I13	PROBLEM	'problem'	I have a problem with this computer.
I14	TELEFONIEREN	'to call'	I am on the phone (literally: "I am calling")
I15	ANGESTELLT	'employed'	She's employed in a large company.
I16	ABER	'but'	My husband is not at home, but he'll be here soon.
I17	PAPIER	'paper'	I need a yellow sheet of paper.
I18	THEMA	'topic'	The professor is giving a talk on the topic politics.
I19	SCHICKEN	'to send'	I send you a message.
I20	ANTWORT	'answer'	I'm waiting for an answer to my question.
I21	LEHREN	'to teach'	She teaches history in school.
I22	EINVERSTANDEN	'agreed'	My teacher agreed that I can take off tomorrow.
I23	VERDIENEN	'to earn'	The mother must work to earn money.
I24	VORSCHLAGEN	'to suggest'	The family suggests different trips to the man.
I25	SAMMELN	'to collect'	The couple collects coins.
I26	SCHLOSS	'castle'	There is a castle in Lenzburg.
I27	BLAU	'blue' (color)	I like the blue car.
I28	FREUND/KOLLEGE	'friend'	Tomorrow, I'll meet with my friend.
I29	SALZ	'salt'	I'm always cooking with too much salt.
I30	IMMER	'always'	The baby drinks always milk.
I31	HUSTEN	'cough'	I have cough.
I32	ERFOLG	'success'	Facebook has a lot of success.
I33	EI	'egg'	Every morning I'm cooking myself an egg.

I34	WASSER	'water'	Children love water.
I35	SORGEN	'to care'	I care about my child.
I36	MONAT	'month'	My brother visits every month a course.
I37	KOPIEREN	'to copy'	The students often need to copy my signing.
I38	GEDULD	'patience'	I have no patience.
I39	UNFALL	'accident'	There was an accident in Zürich yesterday.
I40	ENTSCHEIDEN	'to decide'	I've to decide if I am flying to the United States.
I41	UNTERSCHRIFT	'signature'	The signature is missing on the contract.
I42	VORGESTERN	'day before yesterday'	The day before yesterday, I was in Basel.
I43	VERKAUF	'sale'	The sale of my car will come up soon.
I44	ERKLÄREN	'to explain'	I'm explaining the political situation of Switzerland to the woman.
I45	KINO	'cinema'	Sometimes the movies in the cinema are subtitled.
I46	THEATER	'theatre' (play)	The play (theatre) will perform in Basel.
I47	NOCHMALS	'again'	I'll ask mama again.
I48	WARTEN	'to wait'	I wait until the doctor arrives.
I49	NAME	'name'	Where is your name on the list?
I50	MÖGLICH	'possible'	The weather forecast announces that tomorrow it will snow possibly.
I51	GRUND	'reason'	The reason is that I've already agreed to meet someone.
I52	FRAGEN	'to ask'	You have to ask your dad.
I53	NASS	'wet'	My trousers are wet, because it rained.
I54	GEBURTSTAG	'birthday'	On Sunday is my child's birthday.
I55	KOMMUNIKATION	'communication'	It is impossible to survive without communication.
I56	FRAU	'woman'	He's looking for a woman.
I57	METALL	'metal'	My candleholder is made of metal.
I58	SPRACHE	'language'	I'm learning the German language.
I59	SCHON	'already'	I've been already to the doctor.
I60	SOMMER	'summer'	During the summer, a lot of people go swimming.
I61	FARBE	'color'	I like the color red.
I62	SCHWIERIG	'difficult'	This task is difficult.
I63	CHAOS	'chaos'	There is total chaos at the central station in Zürich.
I64	GEGEN	'against'	The next soccer game is Switzerland against Germany.
I65	WASCHEN	'to wash'	Please wash your cloths!
I66	TRAUM	'dream'	I had a strange dream last night.
I67	STERBEN	'to die'	This person will die.
I68	KOMISCH	'strange'	The woman is wearing strange cloths.
I69	GRUPPE	'group'	This group (of people) is visiting the museum.
I70	GEHT-MICH-NICHTS-AN	'I don't care' (idiomatic sign)	I don't care.
I71	SPORT	'sports'	Sports keep you fit.
I72	SPITAL	'hospital'	The father visits the child in the hospital.
I73	SPIELEN	'to play'	The children are playing.

I74	TAXI	‘taxi’	The taxi arrives.
I75	AUCH	‘too/also,/as well’	I am joining as well.
I76	BESPRECHEN	‘to talk’	Tomorrow we’ll talk about your work.
I77	FAMILIE	‘family’	My family is rich.
I78	JETZT	‘now’	I am hungry now.
I79	WICHTIG	‘important’	It is important that you drink a lot.
I80	FREUDE	‘joy/happiness’	There is a lot of joy/happiness around.
I81	STRENG	‘strict’	My teacher is strict.
I82	KRANK	‘sick’	My son is sick.
I83	VON	‘from’	This man comes from a company.
I84	ABEND	‘evening’	The sun settles in the evening.
I85	UNSICHER	‘uncertain’	I’m uncertain if my boss will still come.
I86	SPIEGEL	‘mirror’	The mirror is broken.
I87	TEXT	‘text’	The text is long.
I88	UNTERSUCHEN	‘to examine/to investigate’	The doctor examines my blood.
I89	ANDERS	‘different’	The soup tastes different today.
I90	DANN	‘then’	I’ll cook first, then we can eat.
I91	AKZEPTIEREN	‘to accept’	The parents accept that their sun is moving out.
I92	JAHR	‘year’	The year passes by fast.
I93	ERINNERN	‘to remember’	I remember my last vacation in Egypt.